

Package ‘bigPint’

April 21, 2018

Version 0.99.0

Title Big multivariate data plotted interactively

Description Methods for visualizing large multivariate datasets using static and interactive scatter-plot matrices, parallel coordinate plots, and litre plots.

Author Lindsay Rutter, Dianne Cook

Maintainer Lindsay Rutter <l.rutter@iastate.edu>

License GPL

Depends R (>= 3.4.1)

Imports data.table (>= 1.10.4),
datasets (>= 3.4.1),
dplyr (>= 0.7.2),
ggplot2 (>= 2.2.1),
hexbin (>= 1.27.1),
Hmisc (>= 4.0.3),
htmlwidgets (>= 0.9),
plotly (>= 4.7.1),
RColorBrewer (>= 1.1.2),
shiny (>= 1.0.5),
shinydashboard (>= 0.6.1),
tidyr (>= 0.7.0),
DESeq2 (>= 1.16.1),
GGally (>= 1.3.2)

VignetteBuilder knitr

Suggests stringr (>= 0.6.2),
knitr (>= 1.13),
roxygen2 (>= 3.0.0)

RoxygenNote 6.0.1

R topics documented:

bigPint	2
degFC	2
degPCP	3
degPI	3
degScatMat	3
degScatMatPoints	4

degVolcano	4
plotClusters	4
plotDEG	5
plotPCPInteractive	7
plotPermutations	7
plotPermutationsD	8
plotScatterInteractive	8
plotScatterInteractiveDash	9
plotScatterStatic	10
selDelIntPCP	11
selDelIntShadePCP	11
selDelPCP	11
selDelShadePCP	12
selPCP	12
soybean_cn	12
soybean_cn_metrics	13
soybean_ir	14
soybean_ir_metrics	14
Index	16

bigPint	bigPint <i>package</i>
---------	------------------------

Description

bigPint R API

Details

See the README on [GitHub](#)

degFC	<i>Superimpose DEGs onto scatterplot matrix fold change</i>
-------	---

Description

Superimpose DEGs onto scatterplot matrix fold change

Usage

```
degFC(data, dataMetrics, threshFC = threshFC, threshVar = threshVar,
      threshVal = threshVal, bluePointSize = bluePointSize,
      redPointSize = redPointSize, greyPointSize = greyPointSize,
      outDir = outDir)
```

degPCP

*Plot DEGs as parallel coordinate plots***Description**

Plot DEGs as parallel coordinate plots

Usage

```
degPCP(data, dataMetrics, threshVar = threshVar, threshVal = threshVal,
       lineList = lineList, lineSize = lineSize, lineColor = lineColor,
       outDir = outDir, fileName = fileName)
```

degPI

*Superimpose DEGs onto scatterplot matrix of prediction intervals***Description**

Superimpose DEGs onto scatterplot matrix of prediction intervals

Usage

```
degPI(data, dataMetrics, threshVar = threshVar, threshVal = threshVal,
      piLevel = piLevel, bluePointSize = bluePointSize,
      redPointSize = redPointSize, greyPointSize = greyPointSize,
      outDir = outDir)
```

degScatMat

*Superimpose DEGs onto scatterplot matrix***Description**

Superimpose DEGs onto scatterplot matrix

Usage

```
degScatMat(data = data, dataMetrics = dataMetrics, pointSize = pointSize,
           xbins = xbins, threshVar = threshVar, threshVal = threshVal,
           outDir = outDir)
```

degScatMatPoints	<i>Superimpose DEGs onto scatterplot matrix</i>
------------------	---

Description

Superimpose DEGs onto scatterplot matrix

Usage

```
degScatMatPoints(data = data, dataMetrics = dataMetrics,
  pointSize = pointSize, degPointColor = degPointColor,
  threshVar = threshVar, threshVal = threshVal, outDir = outDir,
  fileName = fileName)
```

degVolcano	<i>Superimpose DEGs onto volcano plot</i>
------------	---

Description

Superimpose DEGs onto volcano plot

Usage

```
degVolcano(data, dataMetrics, logFC = logFC, PValue = PValue,
  threshVar = threshVar, threshVal = threshVal, xbins = xbins,
  pointSize = pointSize, outDir = outDir)
```

plotClusters	<i>Plot parallel coordinate lines for clusters</i>
--------------	--

Description

Plot parallel coordinate lines for clusters

Usage

```
plotClusters(data, dataMetrics, nC, threshVar = "FDR", threshVal = 0.05,
  topNum = -1, outDir = getwd(), verbose = FALSE)
```

Arguments

data	data frame containing read counts
dataMetrics	data frame containing metrics
nC	the number of clusters
threshVar	the name of column in dataMetrics object that is used to threshold significance (character string; default "FDR")
threshVal	the maximum value of threshVar from which to select genes to cluster (default 0.05)
outDir	output directory to save all images (default current directory)
verbose	in addition to the usual collective printing of all clusters from a given cluster size, print each cluster from each cluster size into separate images and print the associated IDs of each cluster from each cluster size into separate text files (default is FALSE)
threshNum	the number of genes with the lowest threshVar values to select genes to cluster (default is for threshNum to equal -1 and to select clustering genes based on threshVal. If threshNum is changed to a positive value, then threshVal is overridden)

Examples

```
data(soybean_cn)
data(soybean_cn_metrics)
for (nC in c(3,6)){plotClusters(data=soybean_cn, dataMetrics = soybean_cn_metrics, nC=nC)}
```

plotDEG

*Plot differentially expressed genes***Description**

Superimpose differentially expressed genes onto background plot containing all genes.

Usage

```
plotDEG(data = data, dataMetrics = dataMetrics, outDir = getwd(),
  pointSize = 0.5, bluePointSize = 0.1, redPointSize = 0.1,
  greyPointSize = 0.1, lineSize = 0.1, lineColor = "orange",
  degPointColor = "orange", xbins = 10, piLevel = 0.95, threshFC = 3,
  threshOrth = 3, threshVar = "FDR", threshVal = 0.05, lineList = NULL,
  logFC = "logFC", PValue = "PValue", option = "scatterPoints",
  fileName = "")
```

Arguments

data	data frame containing read counts
dataMetrics	data frame containing metrics
outDir	output directory to save all images (default current directory)
pointSize	size of plotted points (default 0.5; used in "scatterHexagon", "scatterPoints", and "volcano")

bluePointSize	size of plotted blue points (default 0.1; used in "scatterFoldChange" and "scatterOrthogonal" and "scatterPrediction")
redPointSize	size of plotted red points (default 0.1; used in "scatterFoldChange" and "scatterOrthogonal" and "scatterPrediction")
greyPointSize	size of plotted grey points (default 0.1; used in "scatterFoldChange" and "scatterOrthogonal" and "scatterPrediction")
lineSize	size of plotted parallel coordinate lines (default 0.1; used in "parallelCoord")
lineColor	color of plotted parallel coordinate lines (default "orange"; used in "parallelCoord")
degPointColor	color of DEGs plotted as points on scatterplot matrix (default "orange; used in "scatterPoints")
xbins	the number of bins partitioning the range of the plot (default 10; used in "scatterHexagon")
piLevel	prediction interval level (between 0 and 1; default 0.95; used in "scatterPrediction")
threshFC	threshold of fold change (default 3; used in "scatterFoldChange")
threshOrth	threshold of orthogonal distance (default 3; used in "scatterOrthogonal")
threshVar	name of column in dataMetrics object that is used to threshold significance (character string; default "FDR"; used in all)
threshVal	maximum value to threshold significance from threshVar object (default 0.05; used in all)
lineList	list of ID values of genes to be drawn from data as parallel coordinate lines. Use this parameter if you have predetermined genes to be drawn. Otherwise, use dataMetrics, threshVar, and threshVal to create genes to be drawn (default NULL; used in "parallelCoord")
logFC	name of column in dataMetrics object that contains log fold change values (character string; default "logFC"; used in "volcano")
PValue	name of column in dataMetrics object that contains p-values (character string; default "PValue"; used in "volcano")
option	the type of plot (can choose from c("parallelCoord", "scatterFoldChange", "scatterHexagon", "scatterOrthogonal", "scatterPoints", "scatterPrediction", "volcano"); default "scatterPoints")
fileName	the name of the output file (default is based on plot option)
saveFile	save file to outDir (default FALSE)

Details

There are seven options:

- "scatterHexagon": Plot DEGs onto a scatterplot matrix of hexagon binning
- "scatterPoints": Plot DEGs onto a scatterplot matrix of points
- "scatterOrthogonal": Plots DEGs onto a scatterplot matrix of orthogonal distance
- "scatterFoldChange": Plots DEGs onto a scatterplot matrix of fold changes
- "scatterPrediction": Plot DEGs onto a scatterplot matrix of prediction intervals
- "parallelCoord": Plots DEGs as parallel coordinate plots on top of boxplot
- "volcano": Plot DEGs onto a volcano plot

Examples

```
data(soybean_cn)
data(soybean_cn_metrics)
plotDEG(soybean_cn, soybean_cn_metrics)
```

plotPCPInteractive	<i>Highlight parallel coordinate plot lines inside selection box</i>
--------------------	--

Description

Highlight parallel coordinate plot lines inside selection box

Usage

```
plotPCPInteractive(pcpDat, option = "deleteInteger")
```

Arguments

pcpDat	the data frame that contains the parallel coordinate plot values
option	the interactivity option ("deleteInteger", "delete", "highlight"); default ("deleteInteger")

Examples

```
set.seed(3)
f = function(){1.3*rnorm(500)}
pcpDat = data.frame(ID = paste0("ID", 1:500), A.1=f(), A.2=f(), A.3=f(), B.1=f(), B.2=f(), B.3=f())
pcpDat$ID = as.character(pcpDat$ID)
plotPCPInteractive(pcpDat = pcpDat)
```

plotPermutations	<i>Replicate line plot linked with parallel coordinate plot</i>
------------------	---

Description

Replicate line plot linked with parallel coordinate plot

Usage

```
plotPermutations(data = data, nPerm = 10, topThresh = 50,
  threshVal = 0.05, option = "none", outDir = getwd())
```

Arguments

data	the data frame that contains the logged read counts for all samples
nPerm	the number of permutations to perform (default is 10)
topThresh	the number of genes with the lowest FDR values to examine from the real data and the permuted data (default is 50)
threshVal	the FDR threshold to count as significant (default is 0.05)
option	procedures to be performed on data after significant calls for visualization purposes (c("none", "log", "standardize"), default is "none")
outDir	output directory to save all images (default current directory)

Examples

```
data(soybean_cn)
data <- soybean_cn
plotPermutations(data, nPerm = 10, topThresh = 30, outDir = getwd())
```

plotPermutationsD	<i>Replicate line plot linked with parallel coordinate plot</i>
-------------------	---

Description

Replicate line plot linked with parallel coordinate plot

Usage

```
plotPermutationsD(data = data, nPerm = 10, topThresh = 50,
  threshVal = 0.05, option = "none", outDir = getwd())
```

Arguments

data	the data frame that contains the logged read counts for all samples
nPerm	the number of permutations to perform (default is 10)
topThresh	the number of genes with the lowest FDR values to examine from the real data and the permuted data (default is 50)
threshVal	the FDR threshold to count as significant (default is 0.05)
option	procedures to be performed on data after significant calls for visualization purposes (c("none", "log", "standardize"), default is "none")
outDir	output directory to save all images (default current directory)

Examples

```
data(soybean_cn)
data <- soybean_cn
plotPermutationsD(data, nPerm = 10, topThresh = 30, outDir = getwd())
```

plotScatterInteractive	<i>Plot interactive scatterplot matrices</i>
------------------------	--

Description

Plot interactive scatterplot matrices.

Usage

```
plotScatterInteractive(data = data, outDir = getwd(), threshOrth = 3,
  xbins = 10, option = "hexagon")
```


Arguments

data	data frame containing read counts
outDir	output directory to save all images (default current directory)
threshOrth	threshold of orthogonal distance (default 3; used in "orthogonal")
xbins	the number of bins partitioning the range of the plot (default 10; used in "hexagon")
option	the type of plot (can choose from c("hexagon", "foldChange", "orthogonal", "prediction"); default "hexagon")

Details

There are four options:

- "hexagon": Plot interactive scatterplot matrix with hexagon binning
- "foldChange": Plot interactive scatterplot matrix with fold change
- "orthogonal": Plot interactive scatterplot matrix with orthogonal distance
- "prediction": Plot interactive scatterplot matrix with prediction interval

Examples

```
data(soybean_cn)
soybean_cn <- soybean_cn[,1:7]
plotScatterInteractive(soybean_cn)
```

plotScatterInteractiveDash

Plot interactive scatterplot matrices

Description

Plot interactive scatterplot matrices.

Usage

```
plotScatterInteractiveDash(data = data, outDir = getwd(), threshOrth = 3,
  xbins = 10, option = "hexagon")
```

Arguments

data	data frame containing read counts
outDir	output directory to save all images (default current directory)
threshOrth	threshold of orthogonal distance (default 3; used in "orthogonal")
xbins	the number of bins partitioning the range of the plot (default 10; used in "hexagon")
option	the type of plot (can choose from c("hexagon", "foldChange", "orthogonal", "prediction"); default "hexagon")

Details

There are four options:

- "hexagon": Plot interactive scatterplot matrix with hexagon binning
- "foldChange": Plot interactive scatterplot matrix with fold change
- "orthogonal": Plot interactive scatterplot matrix with orthogonal distance
- "prediction": Plot interactive scatterplot matrix with prediction interval

Examples

```
data(soybean_cn)
soybean_cn <- soybean_cn[,1:7]
plotScatterInteractiveDash(soybean_cn)
```

plotScatterStatic	<i>Plot static scatterplot matrices</i>
-------------------	---

Description

Plot static scatterplot matrices.

Usage

```
plotScatterStatic(data = data, outDir = getwd(), saveFile = FALSE,
  pointSize = 1, threshFC = 3, threshOrth = 3, piLevel = 0.95,
  xbins = 10, option = "hexagon")
```

Arguments

data	data frame containing read counts
outDir	output directory to save all images (default current directory)
saveFile	save file to outDir (default FALSE)
pointSize	size of plotted points (default 1; used in "foldChange", "orthogonal", "prediction", and "point")
threshFC	threshold for the fold change (default 3; used in "foldChange")
threshOrth	threshold of orthogonal distance (default 3; used in "orthogonal")
piLevel	prediction interval level (between 0 and 1; default 0.95; used in "prediction")
xbins	the number of bins partitioning the range of the plot (default 10; used in "hexagon")
option	the type of plot (can choose from c("hexagon", "foldChange", "orthogonal", "prediction", "point"); default "hexagon")
fileName	the name of the output file (default is based on plot option)

Details

There are five options:

- "hexagon": Plot static scatterplot matrix with hexagon binning
- "foldChange": Plot static scatterplot matrix with fold change
- "orthogonal": Plot static scatterplot matrix with orthogonal distance
- "prediction": Plot static scatterplot matrix with prediction interval
- "point": Plot static scatterplot matrix with raw points

Examples

```
data(soybean_cn)
soybean_cn <- soybean_cn
plotScatterStatic(soybean_cn)
```

selDelIntPCP	<i>Delete parallel coordinate plot lines inside integers of selection box</i>
--------------	---

Description

Delete parallel coordinate plot lines inside integers of selection box

Usage

```
selDelIntPCP(pcpDat)
```

Arguments

pcpDat	the data frame that contains the parallel coordinate plot values
--------	--

selDelIntShadePCP	<i>Delete parallel coordinate plot lines inside integers of shaded selection box</i>
-------------------	--

Description

Delete parallel coordinate plot lines inside integers of shaded selection box

Usage

```
selDelIntShadePCP(pcpDat)
```

Arguments

pcpDat	the data frame that contains the parallel coordinate plot values
--------	--

selDelPCP	<i>Delete parallel coordinate plot lines inside selection box</i>
-----------	---

Description

Delete parallel coordinate plot lines inside selection box

Usage

```
selDelPCP(pcpDat)
```

Arguments

pcpDat	the data frame that contains the parallel coordinate plot values
--------	--

selDelShadePCP	<i>Delete parallel coordinate plot lines inside shaded selection box</i>
----------------	--

Description

Delete parallel coordinate plot lines inside shaded selection box

Usage

```
selDelShadePCP(pcpDat)
```

Arguments

pcpDat	the data frame that contains the parallel coordinate plot values
--------	--

selPCP	<i>Highlight parallel coordinate plot lines inside selection box</i>
--------	--

Description

Highlight parallel coordinate plot lines inside selection box

Usage

```
selPCP(pcpDat)
```

Arguments

pcpDat	the data frame that contains the parallel coordinate plot values
--------	--

soybean_cn	<i>Normalized soybean cotyledon data</i>
------------	--

Description

This dataset contains normalized RNA-sequencing read counts from soybean cotyledon across three time stages of development. Early stage cotyledons were collected four days after planting and were green but closed. Middle stage cotyledons were collected while green and open, soon after the plant generated its first set of unifoliate leaves. Late stage cotyledons were collected immediately after the initiation of yellowing and shrinking.

Usage

```
data(soybean_cn)
```

Format

a RData instance, 1 row per gene

Details

Normalized soybean cotyledon data

- ID gene name
- S1.1 early stage replicate 1 normalized read counts
- S1.2 early stage replicate 2 normalized read counts
- S1.3 early stage replicate 3 normalized read counts
- S2.1 middle stage replicate 1 normalized read counts
- S2.2 middle stage replicate 2 normalized read counts
- S2.3 middle stage replicate 3 normalized read counts
- S3.1 late stage replicate 1 normalized read counts
- S3.2 late stage replicate 2 normalized read counts
- S3.3 late stage replicate 3 normalized read counts

References

Brown AV, Hudson KA (2015) Developmental profiling of gene expression in soybean trifoliolate leaves and cotyledons. BMC Plant Biol 15:169

soybean_cn_metrics	<i>Normalized soybean cotyledon metrics</i>
--------------------	---

Description

This data contains metrics for normalized RNA-sequencing read counts from soybean cotyledon across three time stages of development. Early stage cotyledons were collected four days after planting and were green but closed. Middle stage cotyledons were collected while green and open, soon after the plant generated its first set of unifoliolate leaves. Late stage cotyledons were collected immediately after the initiation of yellowing and shrinking. The metrics include the log fold change, log counts per million, likelihood ratio, p-values, and FDR values for all genes and all pairwise combinations of treatment groups.

Usage

```
data(soybean_cn_metrics)
```

Format

a RData instance, 1 list per treatment group combination and 1 row per gene

Details

Normalized soybean cotyledon metrics

- ID gene name
- logFC log fold change
- logCPM log counts per million
- LR likelihood ratio
- PValue p-value
- FDR FDR value

See Also

[soybean_cn](#) for information about the treatment groups

soybean_ir	<i>Raw soybean leaves data</i>
------------	--------------------------------

Description

This dataset contains raw RNA-sequencing read counts from a soybean dataset that compared leaves that were exposed to iron-rich (iron-positive) soil conditions versus leaves that were exposed to iron-poor (iron-negative) soil conditions. The data was collected 120 minutes after iron conditions were initiated.

Usage

```
data(soybean_ir)
```

Format

a RData instance, 1 row per gene

Details

Raw soybean leaves data

- ID gene name
- N.1 iron-negative condition replicate 1 raw read counts
- N.2 iron-negative condition replicate 2 raw read counts
- N.3 iron-negative condition replicate 3 raw read counts
- P.1 iron-positive condition replicate 1 raw read counts
- P.2 iron-positive condition replicate 2 raw read counts
- P.3 iron-positive condition replicate 3 raw read counts

soybean_ir_metrics	<i>Raw soybean leaves metrics</i>
--------------------	-----------------------------------

Description

This data contains metrics for raw RNA-sequencing read counts from a soybean dataset that compared leaves that were exposed to iron-rich (iron-positive) soil conditions versus leaves that were exposed to iron-poor (iron-negative) soil conditions. The data was collected 120 minutes after iron conditions were initiated. The metrics include the log fold change and the p-values for all genes and all pairwise combinations of treatment groups.

Usage

```
data(soybean_ir_metrics)
```

Format

a RData instance, 1 list per treatment group combination and 1 row per gene

Details

Raw soybean leaves metrics

- ID gene name
- logFC log fold change
- PValue p-value

See Also

[soybean_ir](#) for information about the treatment groups

Index

*Topic **datasets**

- soybean_cn, [12](#)
- soybean_cn_metrics, [13](#)
- soybean_ir, [14](#)
- soybean_ir_metrics, [14](#)

bigPint, [2](#)
bigPint-package (bigPint), [2](#)

degFC, [2](#)
degPCP, [3](#)
degPI, [3](#)
degScatMat, [3](#)
degScatMatPoints, [4](#)
degVolcano, [4](#)

plotClusters, [4](#)
plotDEG, [5](#)
plotPCPInteractive, [7](#)
plotPermutations, [7](#)
plotPermutationsD, [8](#)
plotScatterInteractive, [8](#)
plotScatterInteractiveDash, [9](#)
plotScatterStatic, [10](#)

selDelIntPCP, [11](#)
selDelIntShadePCP, [11](#)
selDelPCP, [11](#)
selDelShadePCP, [12](#)
selPCP, [12](#)
soybean_cn, [12](#), [14](#)
soybean_cn_metrics, [13](#)
soybean_ir, [14](#), [15](#)
soybean_ir_metrics, [14](#)