# Analysis 9 - Getting a closer look at the clusters from the clusters from basic some WT only.

## Purpose

This analysis is based off of `lcmSOM_analysis4_072914.Rmd`, where the original dataset was made. The top 25% of co-efficent of variation.

## Questions

1. What type of gene co-expression patterns are found in this data?
2. Is there any cluster that is enriched in genes from the curated leaf gene list?
3. Are there GO categories that are enriched in the clusters?
4. If there are GO categories, what are the genes that are represented?

## Caveats

1. While the standard deviation of expression of each gene in each tissue were from the reps were calculated, this was not used when creating the self organized map.

## Analysis Set-up

Required Libraries

```
library(VennDiagram)
library(ggplot2)
library(reshape)
library(kohonen)
library(goseq)
library(GO.db)
library(knitr)
source("./clusterFunctions.R")
```

Upload that dataset:

```
genes25 <- read.csv("../data/analysis4.top25.csv")
head(genes25)
```

```
##    X               gene  tf2ambr tf2aother tf2bmbr tf2bother  tf2cmbr
## 1  3 Solyc00g005060.1.1   2.8122 1.388e-17  13.073    0.5228   0.1063
## 2  4 Solyc00g005070.1.1  16.1748 1.420e+01 158.811    4.4798  11.5423
## 3  8 Solyc00g005430.1.1   0.9591 1.563e+00   2.379    0.8714   0.7496
## 4 10 Solyc00g005840.2.1  13.5849 4.424e+01   7.508   19.3277  10.4523
## 5 11 Solyc00g005880.1.1   1.8399 1.124e+00  61.639    2.0358   5.7112
## 6 12 Solyc00g006470.1.1 465.9812 2.906e+02  91.648  289.6609 250.5632
##   tf2cother    wtambr wtaother   wtbmbr  wtbother   wtcmbr wtcother
## 1    0.8377 3.495e-01   0.4027   1.2847 1.388e-17   0.7016    6.511
## 2    3.1077 1.719e+01   4.5236  12.6456 3.462e+00   4.1808   83.519
## 3    0.4416 4.008e-02   0.8329   0.4097 1.358e+00   1.4743    0.946
```

```
## 4   29.7088 1.975e+01  14.0743  11.4830 8.351e+01  15.8112   13.882
## 5    1.7874 1.571e+00   2.1333   2.9694 2.811e+00   1.9558    4.966
## 6  118.9511 2.456e+03 605.8620 108.0947 3.605e+02 499.4475  409.566
##         sd average     cv
## 1   3.8807   2.217 1.7506
## 2  46.7495  27.820 1.6804
## 3   0.6226   1.002 0.6214
## 4  21.3432  23.611 0.9040
## 5  17.0906   7.545 2.2651
## 6 638.1528 495.533 1.2878
```

```
genes25 <- genes25[,c(2,9:14)]
```

```
scale_data <- as.matrix(t(scale(t(genes25[c(2:7)]))))
pca <- prcomp(scale_data, scale=TRUE)

summary(pca)
```

```
## Importance of components:
##                           PC1   PC2   PC3   PC4   PC5      PC6
## Standard deviation     1.325 1.119 1.105 0.970 0.912 2.02e-15
## Proportion of Variance 0.293 0.209 0.204 0.157 0.138 0.00e+00
## Cumulative Proportion  0.293 0.501 0.705 0.862 1.000 1.00e+00
```

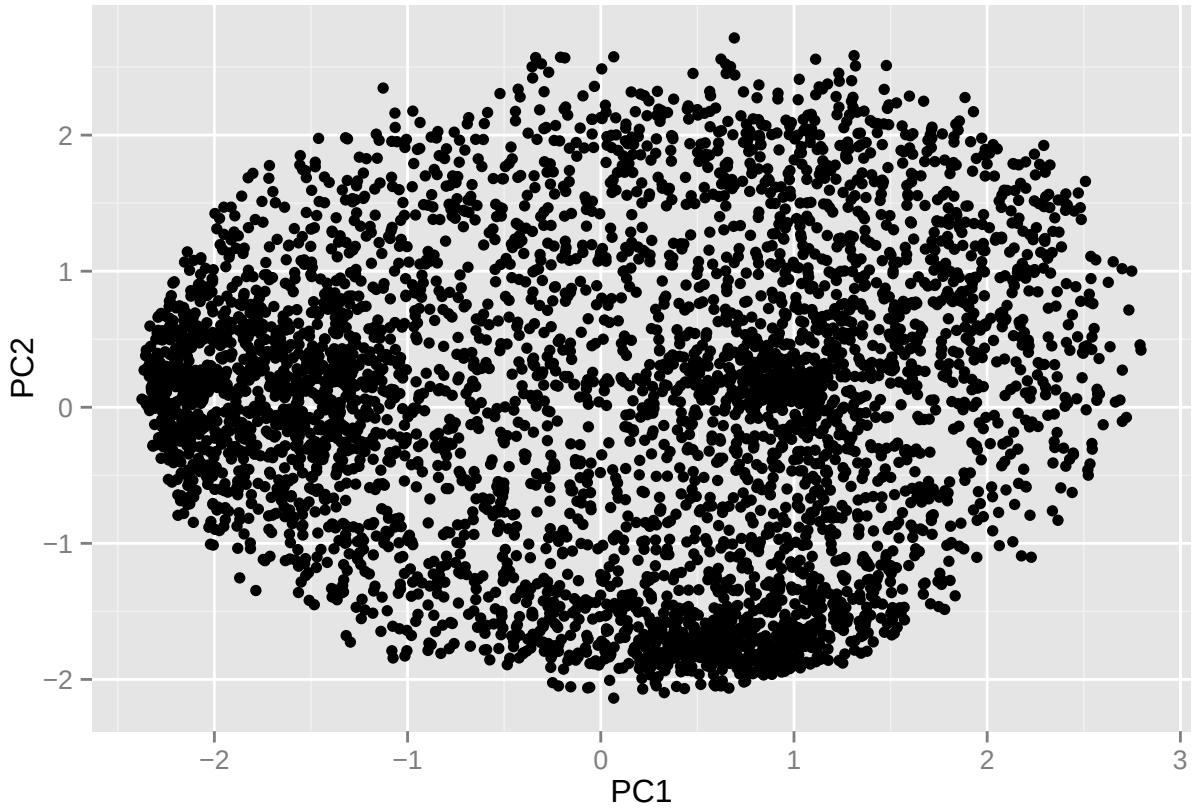
```
pca.scores <- data.frame(pca$x)

data.val <- cbind(genes25, scale_data, pca.scores)
head(data.val)
```

```
##                 gene    wtambr wtaother   wtbmbr  wtbother   wtcmbr
## 1 Solyc00g005060.1.1 3.495e-01   0.4027   1.2847 1.388e-17   0.7016
## 2 Solyc00g005070.1.1 1.719e+01   4.5236  12.6456 3.462e+00   4.1808
## 3 Solyc00g005430.1.1 4.008e-02   0.8329   0.4097 1.358e+00   1.4743
## 4 Solyc00g005840.2.1 1.975e+01  14.0743  11.4830 8.351e+01  15.8112
## 5 Solyc00g005880.1.1 1.571e+00   2.1333   2.9694 2.811e+00   1.9558
## 6 Solyc00g006470.1.1 2.456e+03 605.8620 108.0947 3.605e+02 499.4475
##   wtcother  wtambr wtaother  wtbmbr wtbother  wtcmbr wtcother     PC1
## 1    6.511 -0.4822 -0.46066 -0.1039 -0.62353 -0.3398   2.0100  0.8529
## 2   83.519 -0.1196 -0.52622 -0.2656 -0.56030 -0.5372   2.0089  0.8180
## 3    0.946 -1.4636 -0.01919 -0.7901  0.93674  1.1493   0.1869 -1.0080
## 4   13.882 -0.2374 -0.43921 -0.5314  2.03149 -0.3774  -0.4460 -1.3230
## 5    4.966 -0.9589 -0.49543  0.1936  0.06336 -0.6416   1.8390  0.3961
## 6  409.566  2.0022 -0.15634 -0.7372 -0.44270 -0.2805  -0.3854  0.6217
##       PC2     PC3      PC4      PC5        PC6
## 1 -1.6794 -0.7115 -0.10700  0.01075  2.220e-16
## 2 -1.8799 -0.2741  0.02502 -0.04351 -2.220e-16
## 3  0.8140 -1.3876  1.14932 -0.09713  3.331e-16
## 4  0.2193  0.7110  1.10255  1.45429  1.610e-15
## 5 -1.6693 -0.8672 -0.23448  0.90073  1.887e-15
## 6  0.1852  2.2867  0.39962 -0.66500 -2.942e-15
```

### Visualizing the PCA

```
p <- ggplot(data.val, aes(PC1, PC2))
p + geom_point()
```




### Self Organizing Map - (6,6), large

```
#subset only the scaled gene expression values

head(scale_data)
```

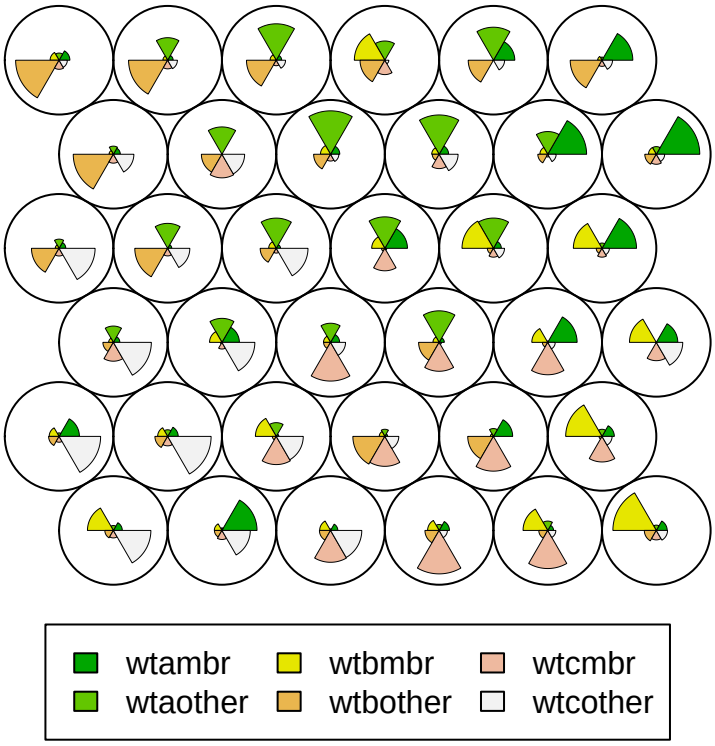
```
##       wtambr wtaother  wtbmbr wtbother  wtcmbr wtcother
## [1,] -0.4822 -0.46066 -0.1039 -0.62353 -0.3398   2.0100
## [2,] -0.1196 -0.52622 -0.2656 -0.56030 -0.5372   2.0089
## [3,] -1.4636 -0.01919 -0.7901  0.93674  1.1493   0.1869
## [4,] -0.2374 -0.43921 -0.5314  2.03149 -0.3774  -0.4460
## [5,] -0.9589 -0.49543  0.1936  0.06336 -0.6416   1.8390
## [6,]  2.0022 -0.15634 -0.7372 -0.44270 -0.2805  -0.3854
```

```
set.seed(6)
som <- som(data=scale_data, somgrid(6,6,"hexagonal")) # This is where you change the size of the map
summary(som)
```

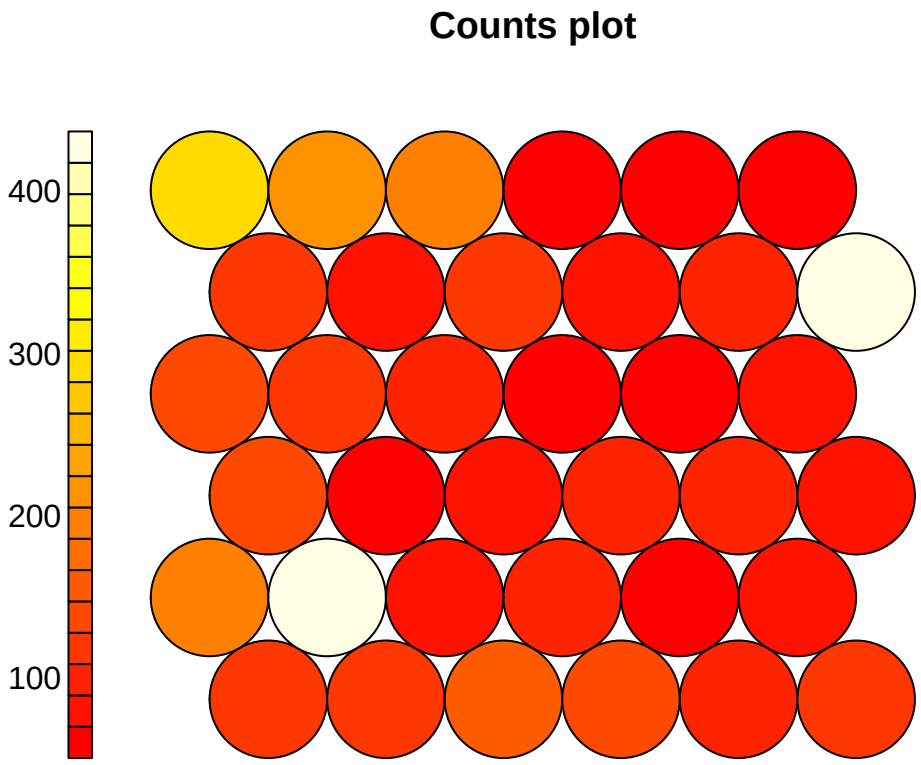
```
## som map of size 6x6 with a hexagonal topology.
## Training data included; dimension is 4618 by 6
## Mean distance to the closest unit in the map: 0.6737
```

### Look at the SOM results

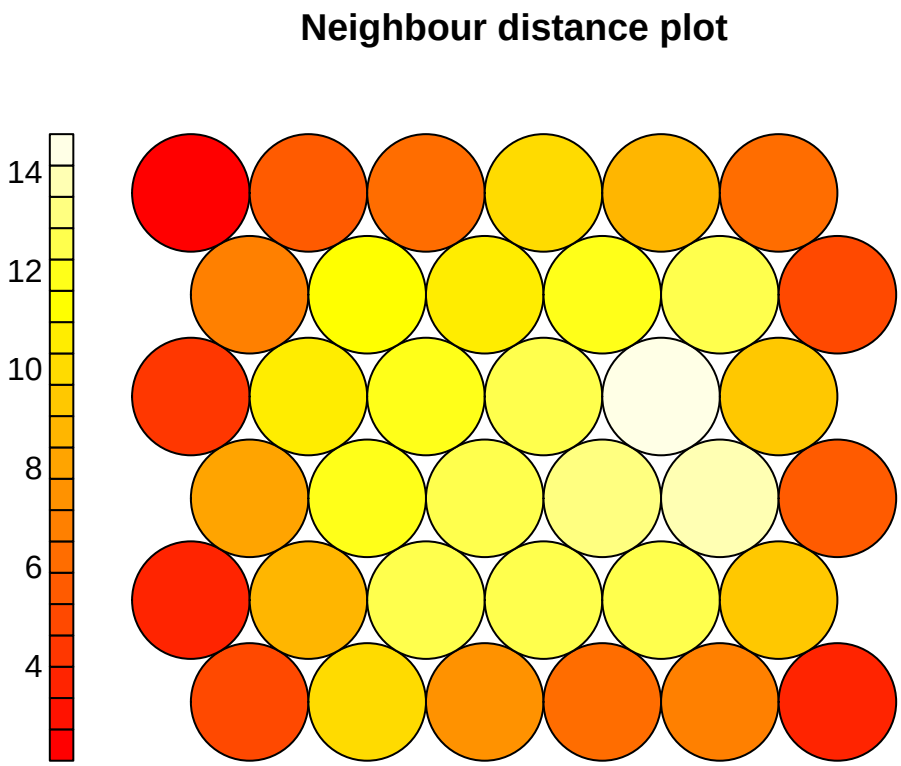
```
plot(som, type = "codes")
```




```
plot(som, type = "counts")
```




```
plot(som, type="dist.neighbours")
```




```
head(som$codes)
```

```
##       wtambr wtaother  wtbmbr wtbother  wtcmbr wtcother
```

```
## [1,] -0.6421  -0.4993  0.6867  -0.7243 -0.4076   1.5866
## [2,]  1.3207  -0.6470 -0.5440  -0.8022 -0.3054   0.9778
## [3,] -0.7651  -0.7102 -0.3252  -0.5241  1.1543   1.1703
## [4,] -0.5398  -0.4355 -0.2816  -0.2866  1.8971  -0.3536
## [5,] -0.8752  -0.2259  0.6039  -0.6945  1.5552  -0.3635
## [6,] -0.4958  -0.4749  1.8671  -0.4565 -0.2819  -0.1580
```

```
som$data <- data.frame(som$data) #changed to dataframe to extract column names easier.

data.val2 <- cbind(data.val,som$unit.classif,som$distances)

#fix to one regex
data.val2$gene <- gsub("^(.*)[.].*", "\\1", data.val2$gene)
data.val2$gene <- gsub("^(.*)[.].*", "\\1", data.val2$gene)

head(data.val2)
```

```
##             gene    wtambr wtaother   wtbmbr  wtbother   wtcmbr wtcother
## 1 Solyc00g005060 3.495e-01   0.4027   1.2847 1.388e-17   0.7016    6.511
## 2 Solyc00g005070 1.719e+01   4.5236  12.6456 3.462e+00   4.1808   83.519
## 3 Solyc00g005430 4.008e-02   0.8329   0.4097 1.358e+00   1.4743    0.946
## 4 Solyc00g005840 1.975e+01  14.0743  11.4830 8.351e+01  15.8112   13.882
## 5 Solyc00g005880 1.571e+00   2.1333   2.9694 2.811e+00   1.9558    4.966
## 6 Solyc00g006470 2.456e+03 605.8620 108.0947 3.605e+02 499.4475  409.566
##    wtambr wtaother  wtbmbr wtbother  wtcmbr wtcother     PC1     PC2
## 1 -0.4822 -0.46066 -0.1039 -0.62353 -0.3398   2.0100  0.8529 -1.6794
## 2 -0.1196 -0.52622 -0.2656 -0.56030 -0.5372   2.0089  0.8180 -1.8799
## 3 -1.4636 -0.01919 -0.7901  0.93674  1.1493   0.1869 -1.0080  0.8140
## 4 -0.2374 -0.43921 -0.5314  2.03149 -0.3774  -0.4460 -1.3230  0.2193
## 5 -0.9589 -0.49543  0.1936  0.06336 -0.6416   1.8390  0.3961 -1.6693
## 6  2.0022 -0.15634 -0.7372 -0.44270 -0.2805  -0.3854  0.6217  0.1852
##       PC3      PC4      PC5        PC6 som$unit.classif som$distances
## 1 -0.7115 -0.10700  0.01075  2.220e-16                8       0.13932
## 2 -0.2741  0.02502 -0.04351 -2.220e-16                7       0.15509
## 3 -1.3876  1.14932 -0.09713  3.331e-16               10       0.32634
## 4  0.7110  1.10255  1.45429  1.610e-15               31       0.09767
## 5 -0.8672 -0.23448  0.90073  1.887e-15                8       0.89742
## 6  2.2867  0.39962 -0.66500 -2.942e-15               30       0.08093
```

Upload the gene expression list.

```
geneList1 <- read.csv("../../../06diffGeneExp/analysis/indvGenes/yasuCuratedGenes/pnas.1402835111.sd06.

#isolate the genes
genesOfInterest <- geneList1[,c(1,3)]
colnames(genesOfInterest) <- c("gene", "name")
names(genesOfInterest) #check
```

```
## [1] "gene" "name"
```

```
#This is a ridiculas around assigning if a gene is a curated gene!
#Figure out more elegent way.
```

```
data.val2$curated <- match(data.val2$gene, genesOfInterest$gene)
data.val2$curated <- gsub("[[:digit:]]+", "yes", data.val2$curated)
data.val2$curated[is.na(data.val2$curated)] <- "no"
```

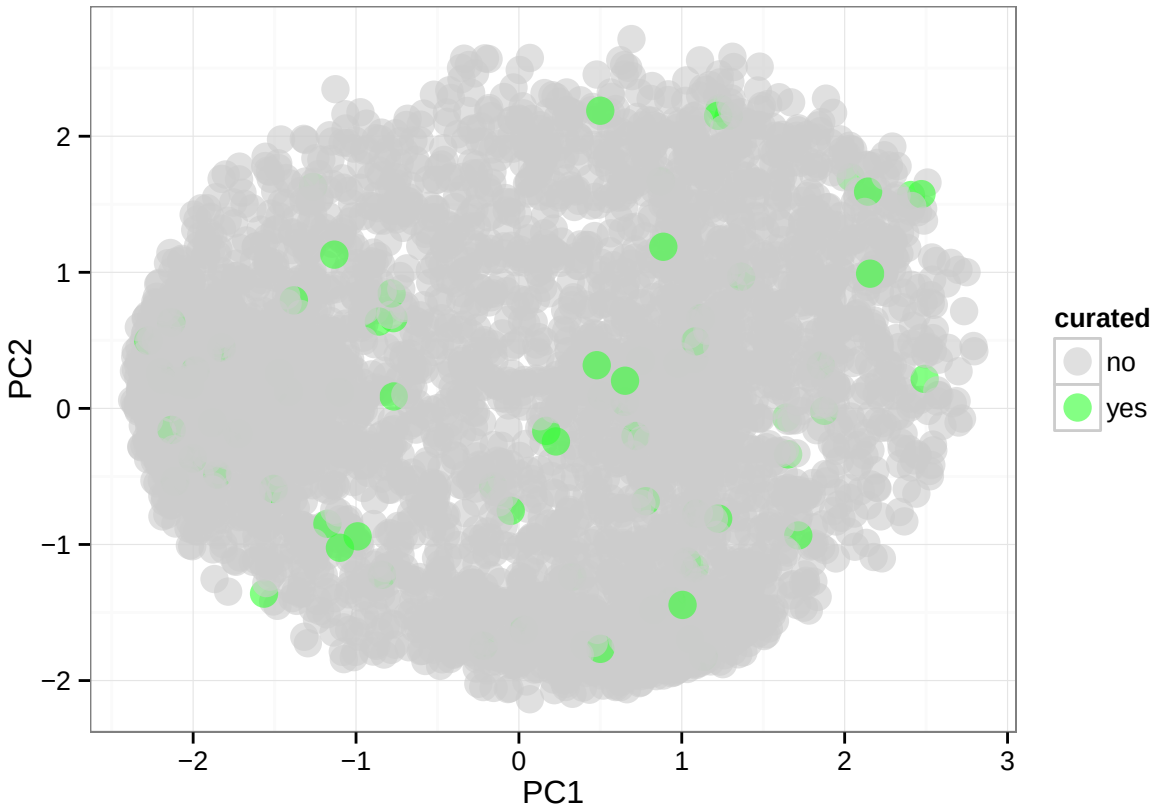
Visualize the major clusters. Here are the leaf curated genes in PC space.

```
head(data.val2)
```

```
##             gene    wtambr wtaother   wtbmbr  wtbother   wtcmbr wtcother
## 1 Solyc00g005060 3.495e-01   0.4027   1.2847 1.388e-17   0.7016    6.511
## 2 Solyc00g005070 1.719e+01   4.5236  12.6456 3.462e+00   4.1808   83.519
## 3 Solyc00g005430 4.008e-02   0.8329   0.4097 1.358e+00   1.4743    0.946
## 4 Solyc00g005840 1.975e+01  14.0743  11.4830 8.351e+01  15.8112   13.882
## 5 Solyc00g005880 1.571e+00   2.1333   2.9694 2.811e+00   1.9558    4.966
## 6 Solyc00g006470 2.456e+03 605.8620 108.0947 3.605e+02 499.4475  409.566
##    wtambr wtaother  wtbmbr wtbother  wtcmbr wtcother     PC1     PC2
## 1 -0.4822 -0.46066 -0.1039 -0.62353 -0.3398   2.0100  0.8529 -1.6794
## 2 -0.1196 -0.52622 -0.2656 -0.56030 -0.5372   2.0089  0.8180 -1.8799
## 3 -1.4636 -0.01919 -0.7901  0.93674  1.1493   0.1869 -1.0080  0.8140
## 4 -0.2374 -0.43921 -0.5314  2.03149 -0.3774  -0.4460 -1.3230  0.2193
## 5 -0.9589 -0.49543  0.1936  0.06336 -0.6416   1.8390  0.3961 -1.6693
## 6  2.0022 -0.15634 -0.7372 -0.44270 -0.2805  -0.3854  0.6217  0.1852
##       PC3      PC4      PC5        PC6 som$unit.classif som$distances
## 1 -0.7115 -0.10700  0.01075  2.220e-16                8       0.13932
## 2 -0.2741  0.02502 -0.04351 -2.220e-16                7       0.15509
## 3 -1.3876  1.14932 -0.09713  3.331e-16               10       0.32634
## 4  0.7110  1.10255  1.45429  1.610e-15               31       0.09767
## 5 -0.8672 -0.23448  0.90073  1.887e-15                8       0.89742
## 6  2.2867  0.39962 -0.66500 -2.942e-15               30       0.08093
##   curated
## 1      no
## 2      no
## 3      no
## 4      no
## 5      no
## 6      no
```

```
p <- ggplot(data.val2, aes(PC1, PC2, color = curated))

p + geom_point(size=I(5), alpha = 0.6) +
  scale_colour_manual(values=c("#cccccc", "#33ff33")) +
  theme_bw()
```




### Visualize by Cluster

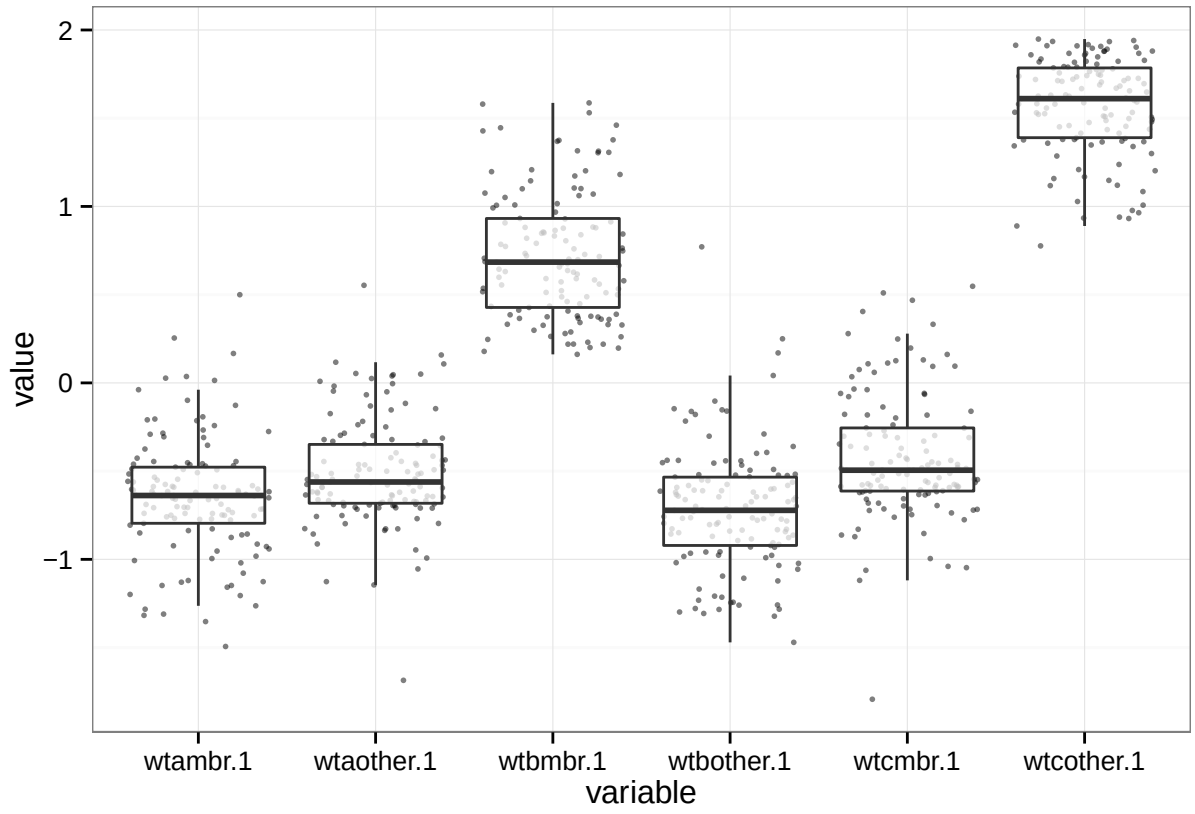
Read in data used for GO enrichment analysis.

```
geneLength <- read.csv("../../../07GO_enrichment/requisiteData/normalized_genes_length.csv")
cate <- read.table("../../../07GO_enrichment/requisiteData/melted.GOTable.txt",header=TRUE)
```

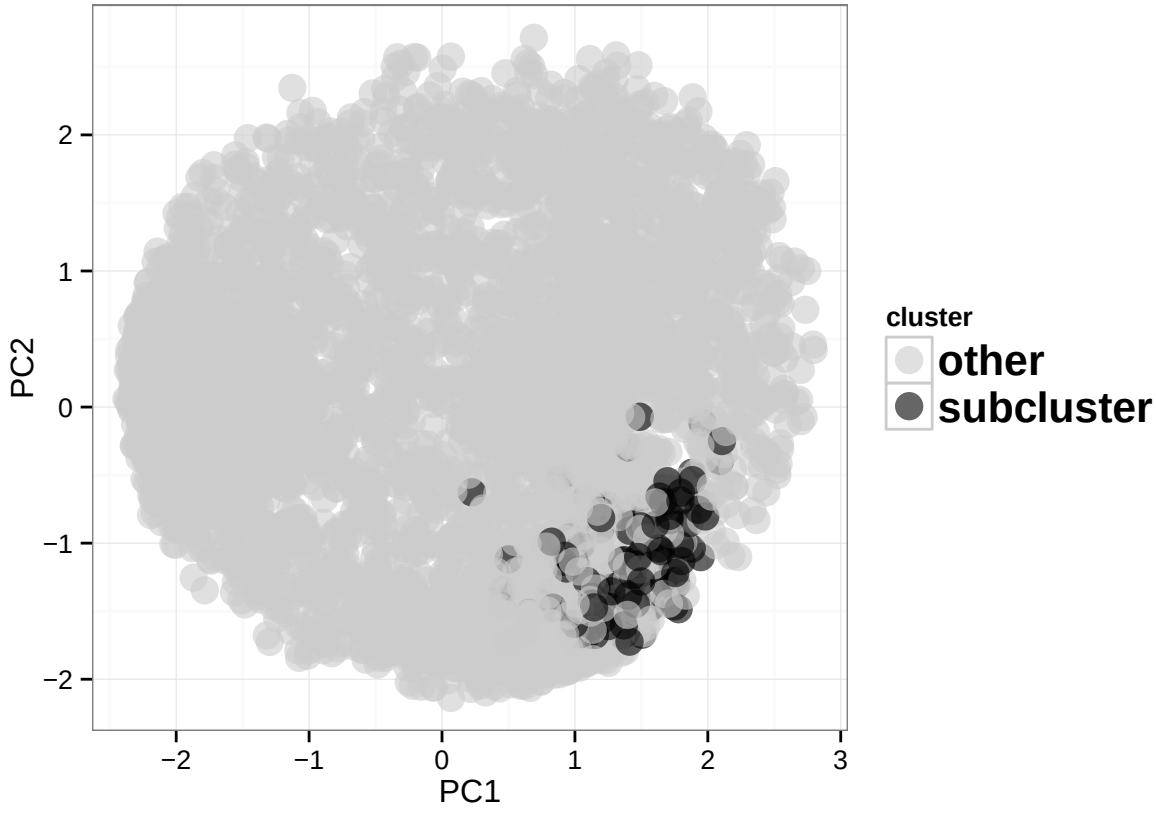
### Cluster 1

```
clusterVis(1)
```

```
## Using gene as id variables
```

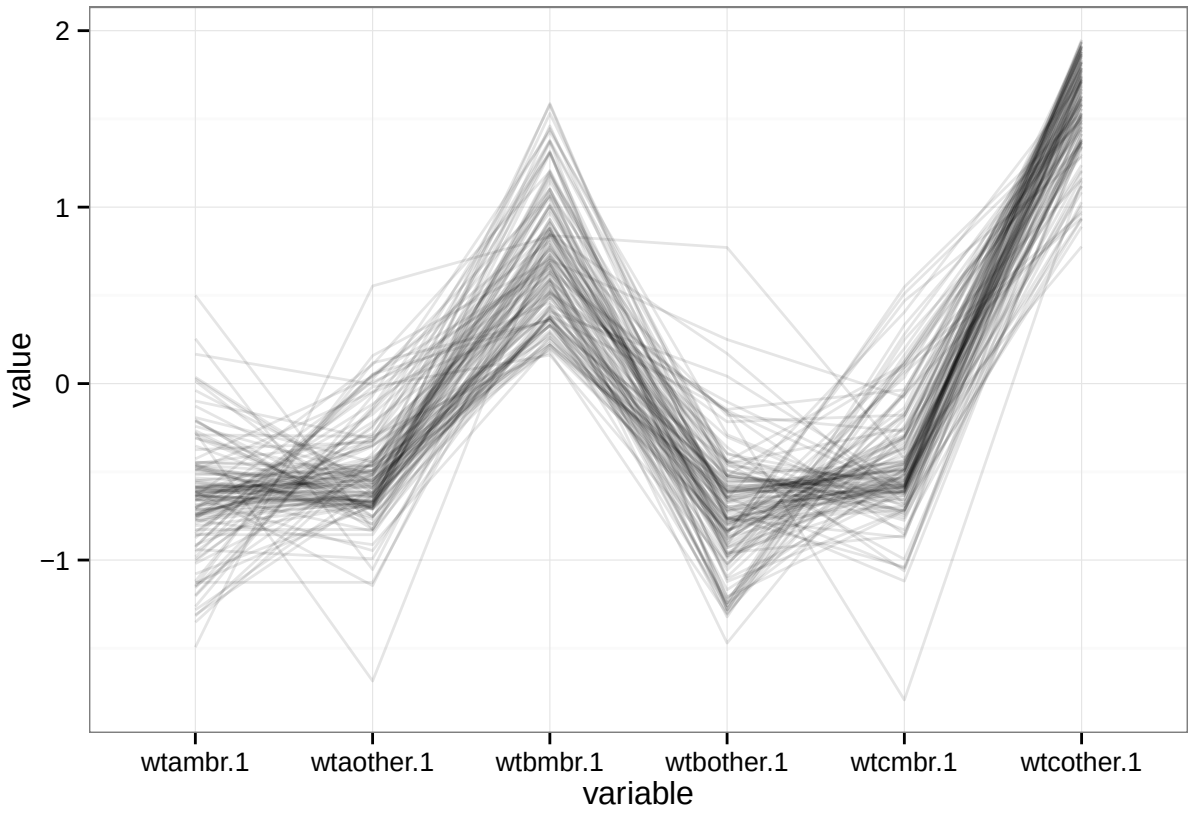



```
clusterVis_PCA(1)
```



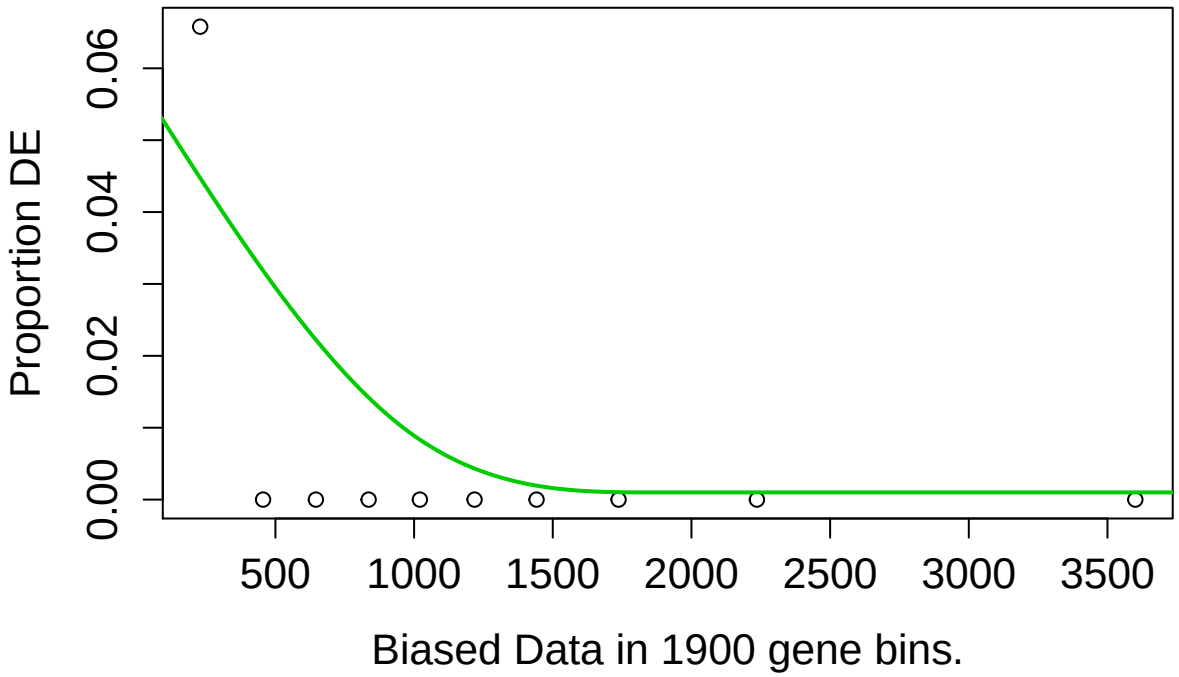
```
clusterVis_line(1)
```

```
## Using gene, curated as id variables
```




```
clusterGO(1)
```

```
## Using manually entered categories.
## For 3061 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```




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

```
y <- genesInClust(1, data.val2, annotation)
```

```
## [1] 125
```

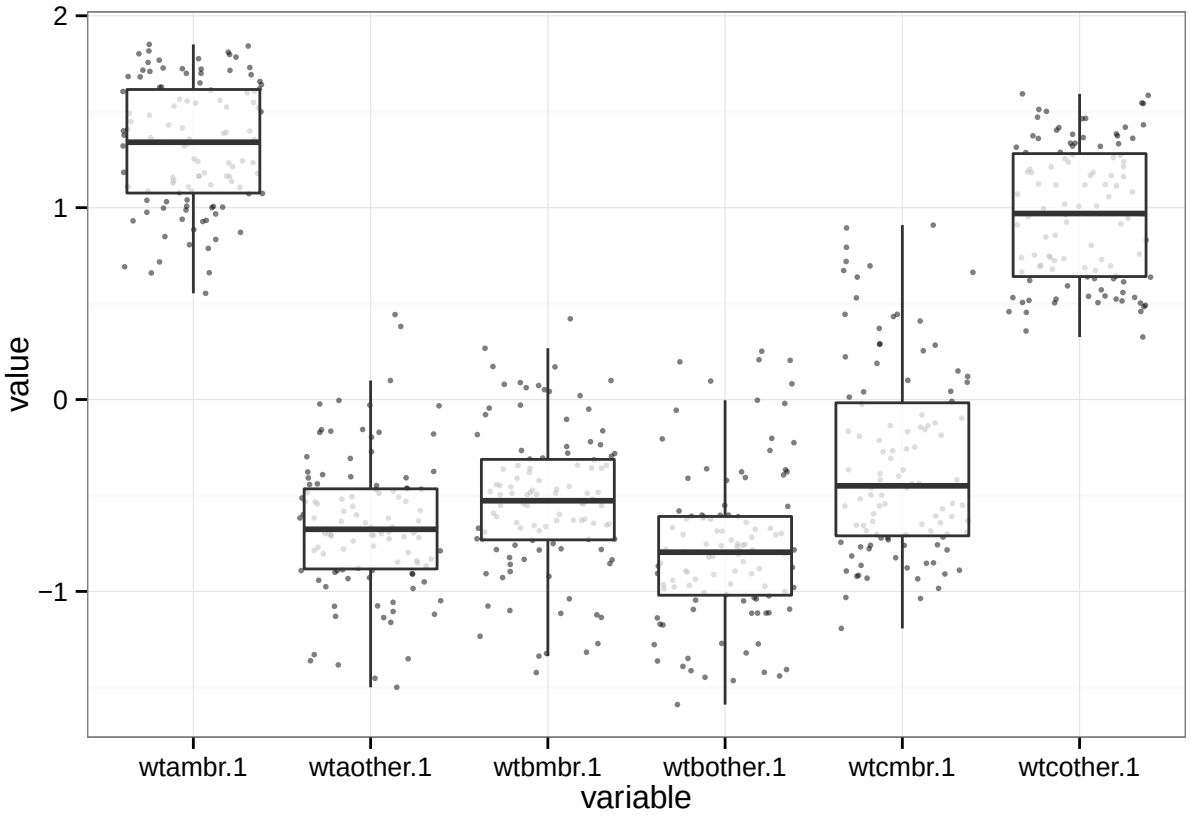
```
#If value intersects with a leaf curated gene
intersect(y$ITAG, genesOfInterest$genes)
```

```
## NULL
```

**Cluster 2**

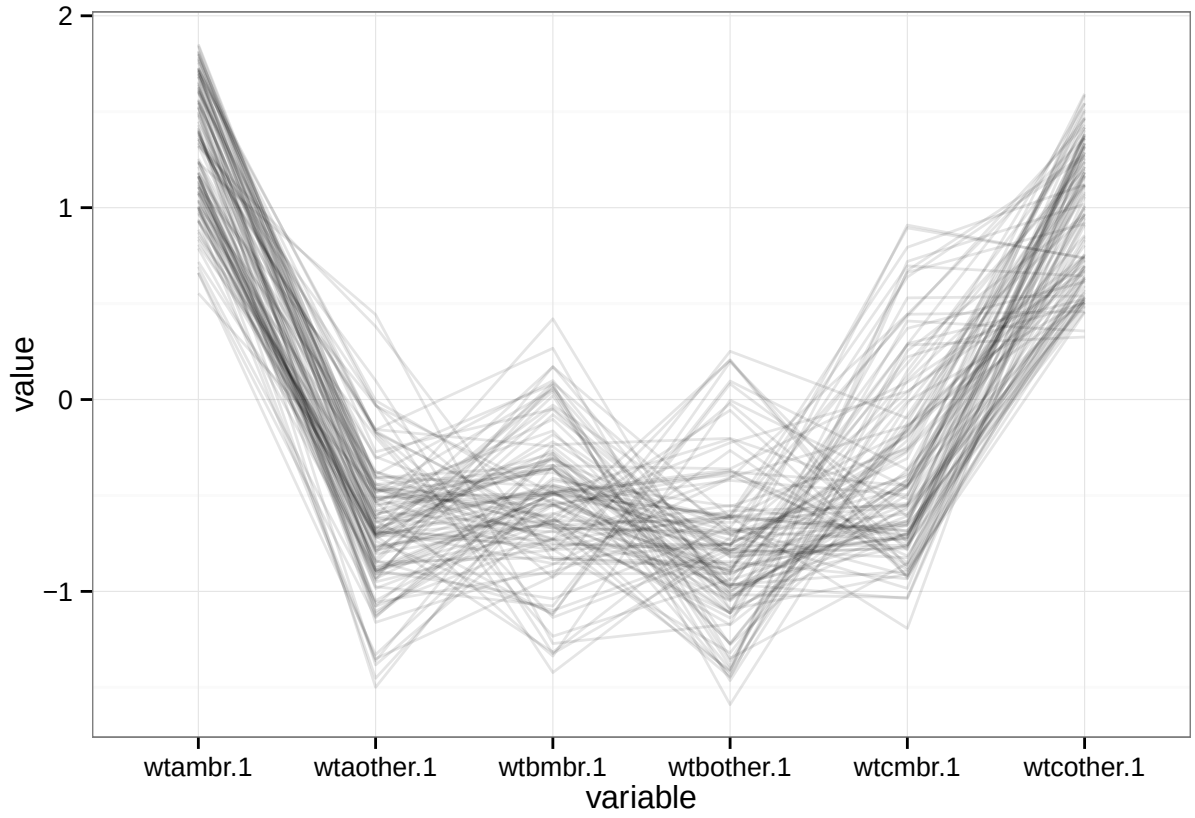
```
clusterVis(2)
```

```
## Using gene as id variables
```

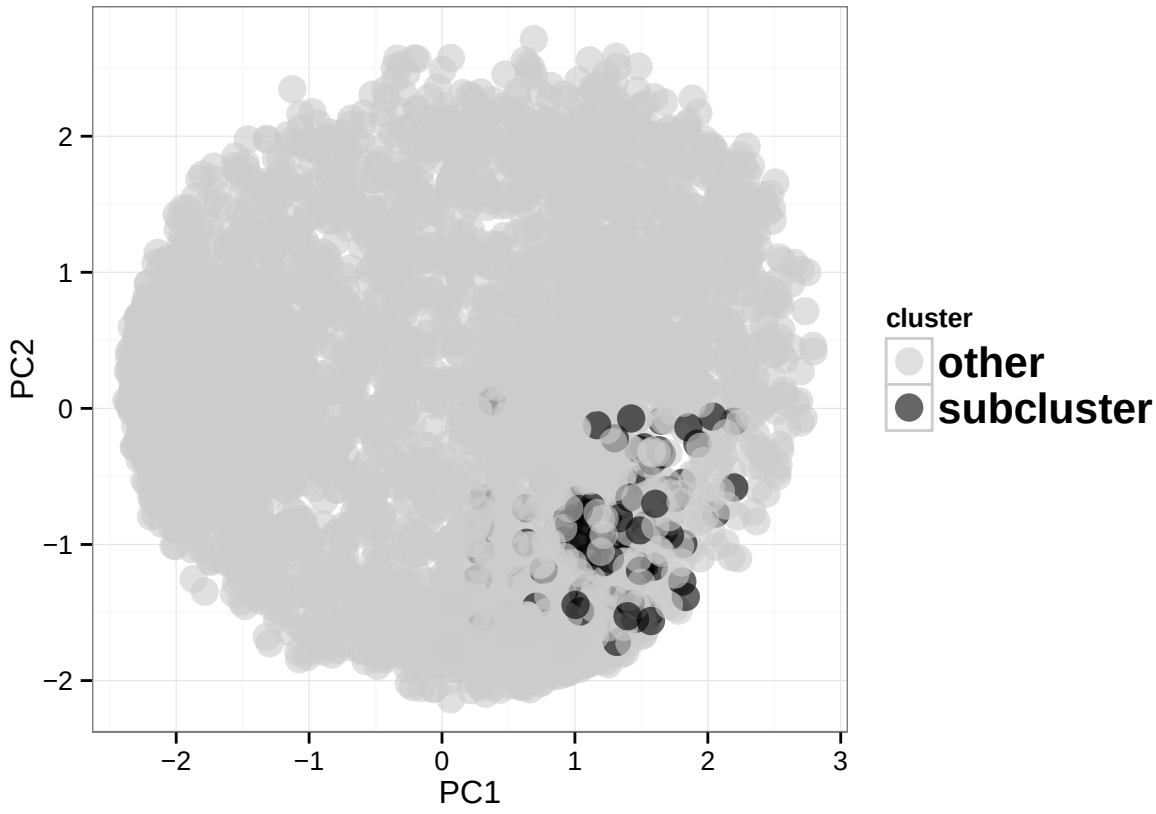



```
clusterVis_line(2)
```

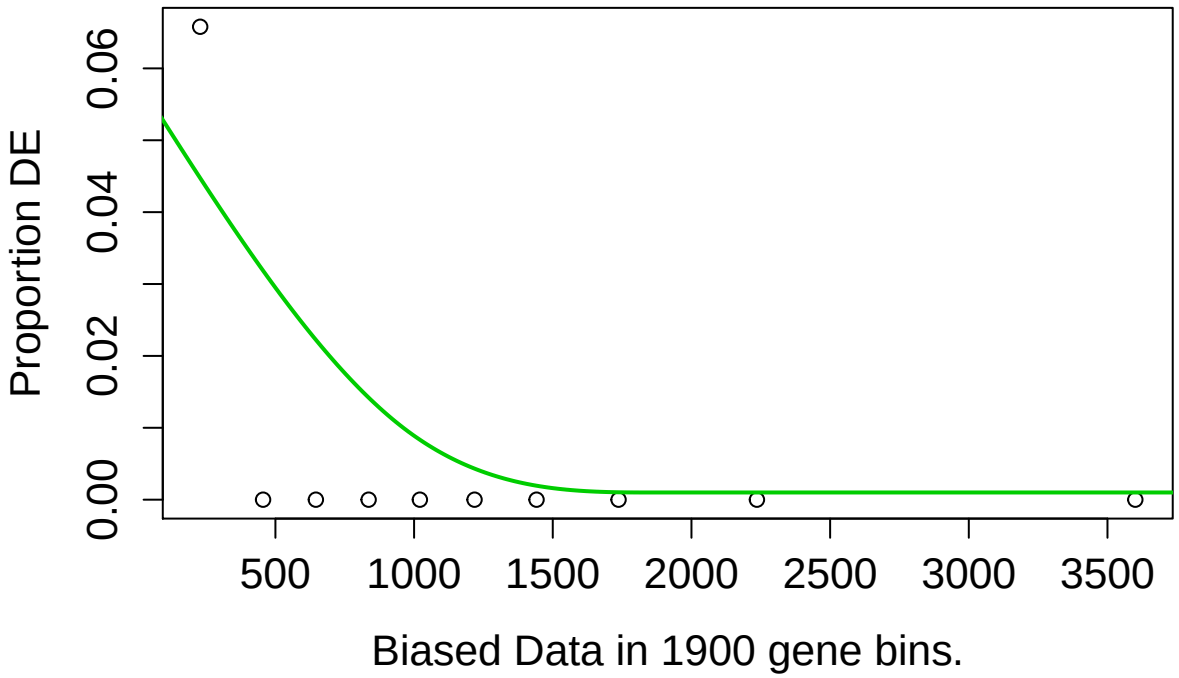
```
## Using gene, curated as id variables
```



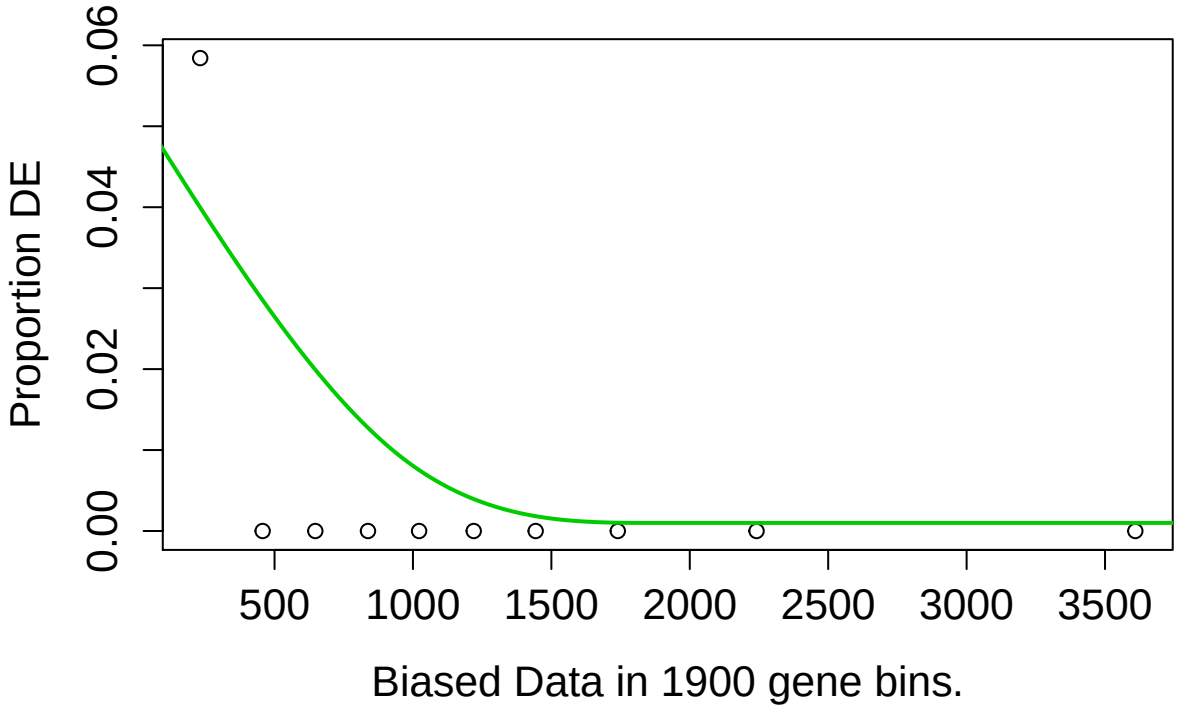

```
clusterVis_PCA(2)
```



```
clusterGO(2)
```




```
## Using manually entered categories.
## For 3047 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```



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

```
y <- genesInClust(20, data.val2, annotation)
```

```
## [1] 125
```

```
kable(y)
```

```
##
##
## |ITAG           |AGI       |gene_name
## |:--------------|:---------|:-----------------------------------------------------------------------
## |Solyc00g014800 |NA        |NA
## |Solyc00g027120 |NA        |NA
## |Solyc00g036520 |NA        |NA
## |Solyc00g052540 |NA        |NA
## |Solyc00g166690 |NA        |NA
## |Solyc00g313030 |NA        |NA
## |Solyc01g010210 |NA        |NA
## |Solyc01g016630 |NA        |NA
## |Solyc01g056770 |NA        |NA
## |Solyc01g057530 |NA        |NA
## |Solyc01g058430 |NA        |NA
## |Solyc01g066510 |AT5G49150 |Encodes a transmembrane domain containing protein expressed in sperm cel
## |Solyc01g081020 |AT4G37380 |pentatricopeptide (PPR) repeat-containing protein; similar to pentatrico
## |Solyc01g081290 |NA        |NA
## |Solyc01g097380 |AT1G60680 |A member of ARF GAP domain (AGD), A thaliana has 15 members, grouped int
## |Solyc01g097680 |NA        |NA
## |Solyc01g097690 |NA        |NA
## |Solyc01g106670 |NA        |NA
## |Solyc01g107960 |NA        |NA
## |Solyc01g109420 |AT4G36750 |quinone reductase family protein; similar to quinone reductase family pr
## |Solyc01g111480 |NA        |NA
## |Solyc01g111770 |AT4G38500 |similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G28240.1); sim
## |Solyc02g025080 |NA        |NA
## |Solyc02g055330 |NA        |NA
## |Solyc02g063170 |NA        |NA
## |Solyc02g065520 |NA        |NA
## |Solyc02g067200 |NA        |NA
## |Solyc02g068840 |NA        |NA
## |Solyc02g069170 |NA        |NA
## |Solyc02g080350 |AT1G77170 |pentatricopeptide (PPR) repeat-containing protein; similar to binding [A
## |Solyc02g081130 |AT4G22140 |DNA binding; similar to bromo-adjacent homology (BAH) domain-containing
## |Solyc02g090730 |AT1G14870 |similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G35525.1); sim
## |Solyc03g044030 |NA        |NA
## |Solyc03g044360 |NA        |NA
## |Solyc03g053030 |NA        |NA
## |Solyc03g062860 |NA        |NA
## |Solyc03g064010 |AT3G57830 |leucine-rich repeat transmembrane protein kinase, putative; similar to l
## |Solyc03g078740 |NA        |NA
## |Solyc03g082420 |AT4G27670 |chloroplast located small heat shock protein.
## |Solyc03g082770 |NA        |NA
## |Solyc03g083740 |NA        |NA
```

```
## |Solyc03g093370 |NA         |NA
## |Solyc03g114570 |NA         |NA
## |Solyc03g120950 |NA         |NA
## |Solyc03g123540 |AT1G54050 |17.4 kDa class III heat shock protein (HSP17.4-CIII); similar to AT-HSP1
## |Solyc04g005770 |AT2G02750 |pentatricopeptide (PPR) repeat-containing protein; similar to pentatricop
## |Solyc04g014940 |NA         |NA
## |Solyc04g015310 |NA         |NA
## |Solyc04g039710 |NA         |NA
## |Solyc04g076690 |NA         |NA
## |Solyc04g078100 |NA         |NA
## |Solyc04g080170 |AT1G45160 |kinase; similar to protein kinase, putative [Arabidopsis thaliana] (TAIR
## |Solyc05g009970 |NA         |NA
## |Solyc05g010080 |NA         |NA
## |Solyc05g017870 |NA         |NA
## |Solyc05g039950 |NA         |NA
## |Solyc05g041600 |NA         |NA
## |Solyc05g050550 |NA         |NA
## |Solyc06g008800 |NA         |NA
## |Solyc06g036290 |AT5G52640 |Arabidopsis thaliana 81 kDA heat shock protein. Sequence analysis reveal
## |Solyc06g068700 |AT5G61790 |calnexin 1 (CNX1); Identical to Calnexin homolog 1 precursor (CNX1) [Aral
## |Solyc06g071380 |NA         |NA
## |Solyc06g073850 |NA         |NA
## |Solyc06g076420 |NA         |NA
## |Solyc06g076590 |NA         |NA
## |Solyc06g076810 |NA         |NA
## |Solyc07g015770 |NA         |NA
## |Solyc07g017550 |NA         |NA
## |Solyc07g017830 |NA         |NA
## |Solyc07g020780 |NA         |NA
## |Solyc07g021560 |NA         |NA
## |Solyc07g026760 |NA         |NA
## |Solyc07g032820 |NA         |NA
## |Solyc07g052870 |NA         |NA
## |Solyc07g054790 |NA         |NA
## |Solyc07g055170 |NA         |NA
## |Solyc07g063020 |NA         |NA
## |Solyc07g064690 |AT1G29040 |similar to unknown protein [Oryza sativa (japonica cultivar-group)] (GB:
## |Solyc07g065060 |NA         |NA
## |Solyc07g065720 |NA         |NA
## |Solyc07g066500 |NA         |NA
## |Solyc08g007420 |AT4G17260 |L-lactate dehydrogenase, putative; similar to PMDH1 (PEROXISOMAL NAD-MALA
## |Solyc08g014390 |NA         |NA
## |Solyc08g023640 |AT4G21065 |binding; similar to pentatricopeptide (PPR) repeat-containing protein [A
## |Solyc08g075210 |NA         |NA
## |Solyc08g076010 |NA         |NA
## |Solyc08g076330 |NA         |NA
## |Solyc08g078080 |NA         |NA
## |Solyc09g008110 |NA         |NA
## |Solyc09g009140 |AT5G06680 |Encodes protein similar to yeast SCP98. Yeast SCP98 is essential for the
## |Solyc09g014620 |NA         |NA
## |Solyc09g042380 |NA         |NA
## |Solyc09g059490 |NA         |NA
## |Solyc09g075170 |AT3G62890 |binding; similar to binding [Arabidopsis thaliana] (TAIR:AT5G40405.1); s
## |Solyc09g082690 |AT3G22840 |Encodes an early light-inducible protein.
```

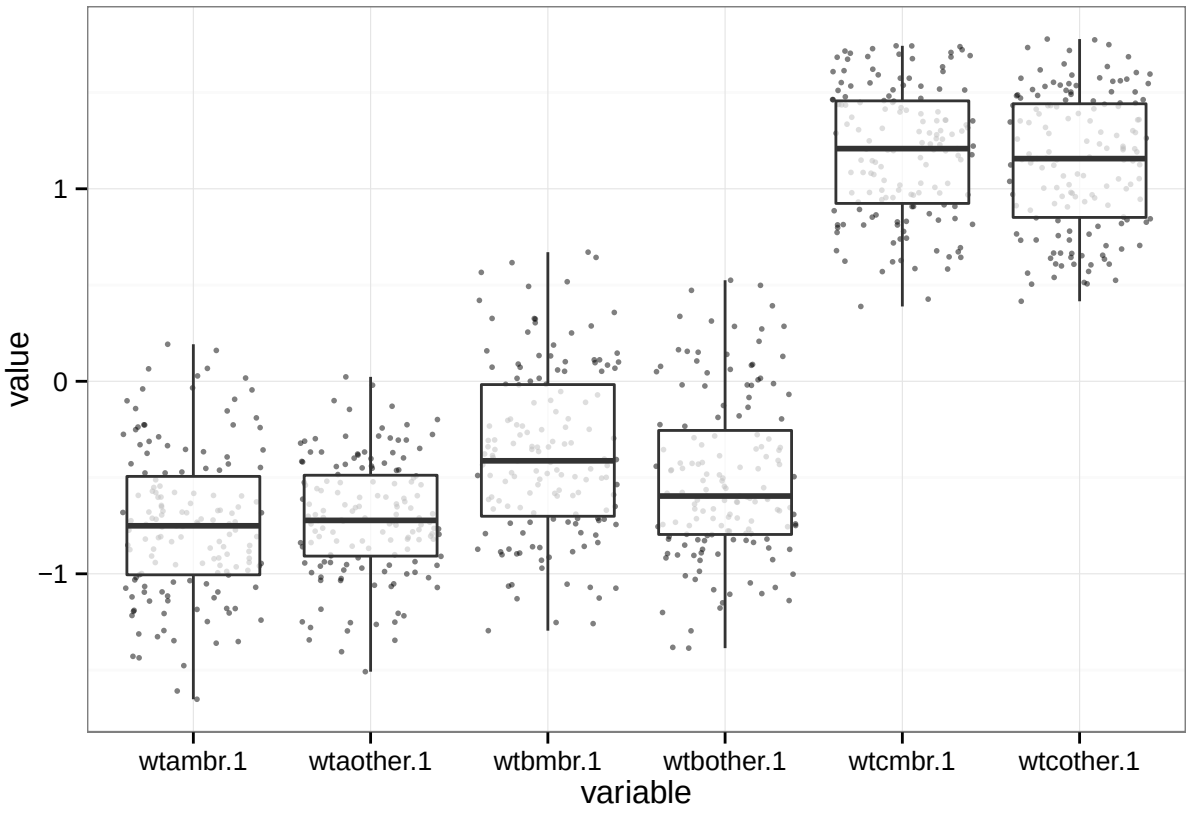
```
## |Solyc10g008880 |AT1G28520 |similar to unknown protein [Arabidopsis thaliana] (TAIR:AT2G42400.1); sim
## |Solyc10g038050 |NA        |NA
## |Solyc10g038070 |NA        |NA
## |Solyc10g039300 |NA        |NA
## |Solyc10g049740 |AT1G05120 |SNF2 domain-containing protein / helicase domain-containing protein / RIN
## |Solyc10g049960 |NA        |NA
## |Solyc10g054680 |NA        |NA
## |Solyc10g078800 |NA        |NA
## |Solyc10g081570 |NA        |NA
## |Solyc11g005240 |NA        |NA
## |Solyc11g008020 |AT5G03530 |ATRAB ALPHA (Arabidopsis Rab GTPase homolog C2a); GTP binding; similar to
## |Solyc11g008060 |NA        |NA
## |Solyc11g008380 |AT3G10370 |mitochondrial FAD-dependent glycerol-3-phosphate dehydrogenase. possibly
## |Solyc11g010220 |NA        |NA
## |Solyc11g011340 |AT4G39330 |mannitol dehydrogenase, putative; Identical to Probable mannitol dehydrog
## |Solyc11g030910 |NA        |NA
## |Solyc11g044680 |NA        |NA
## |Solyc11g065020 |NA        |NA
## |Solyc11g065650 |NA        |NA
## |Solyc11g071370 |AT1G05750 |PDE247 (PIGMENT DEFECTIVE 247); binding; similar to pentatricopeptide (PP
## |Solyc11g071740 |NA        |NA
## |Solyc11g071760 |NA        |NA
## |Solyc12g008610 |NA        |NA
## |Solyc12g010420 |NA        |NA
## |Solyc12g036870 |NA        |NA
## |Solyc12g040330 |NA        |NA
## |Solyc12g042630 |NA        |NA
## |Solyc12g056200 |AT4G11350 |similar to fringe-related protein [Arabidopsis thaliana] (TAIR:AT4G23490
## |Solyc12g056890 |NA        |NA
## |Solyc12g095870 |AT4G32830 |Encodes a member of a family of Ser/Thr kinases whose activities peak dur
```

### Cluster 3

Ambr in WT is higher than in *tf2* and the WT genes in this cluster have a tight pattern.

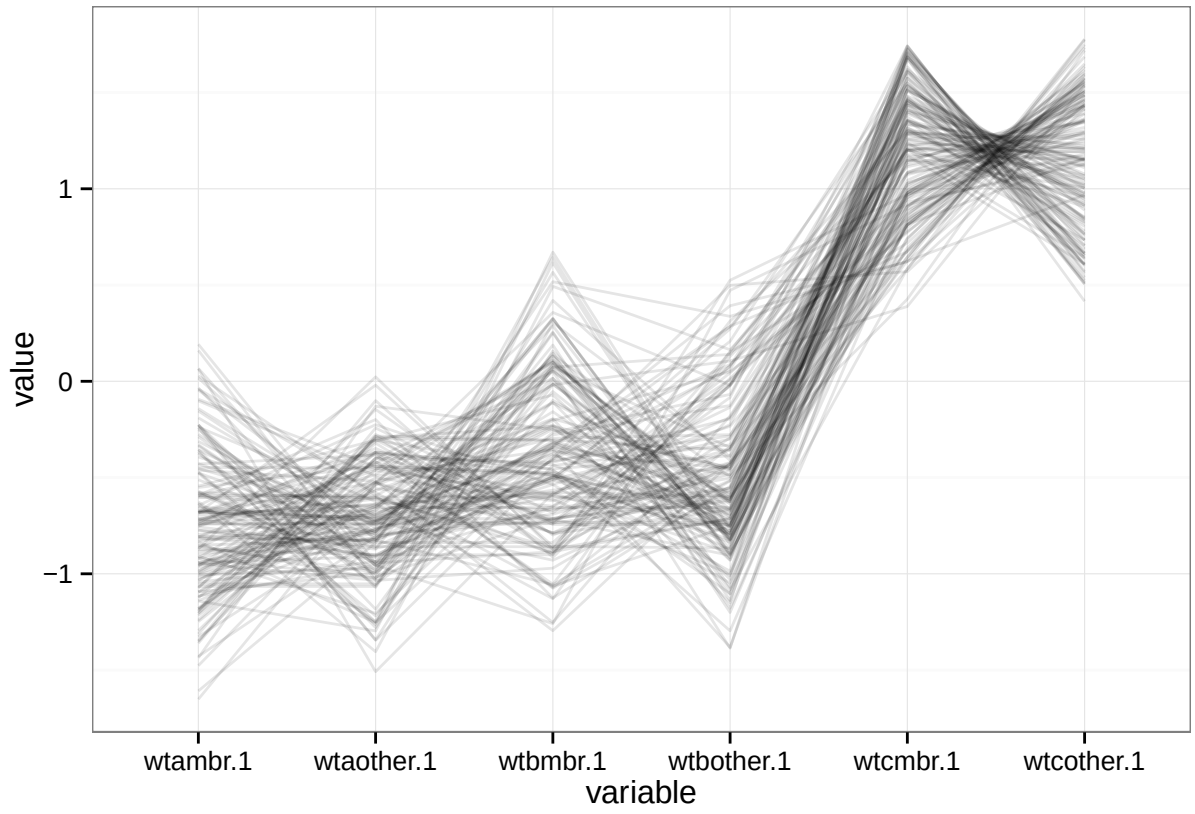
```
clusterVis(3)
```

```
## Using gene as id variables
```

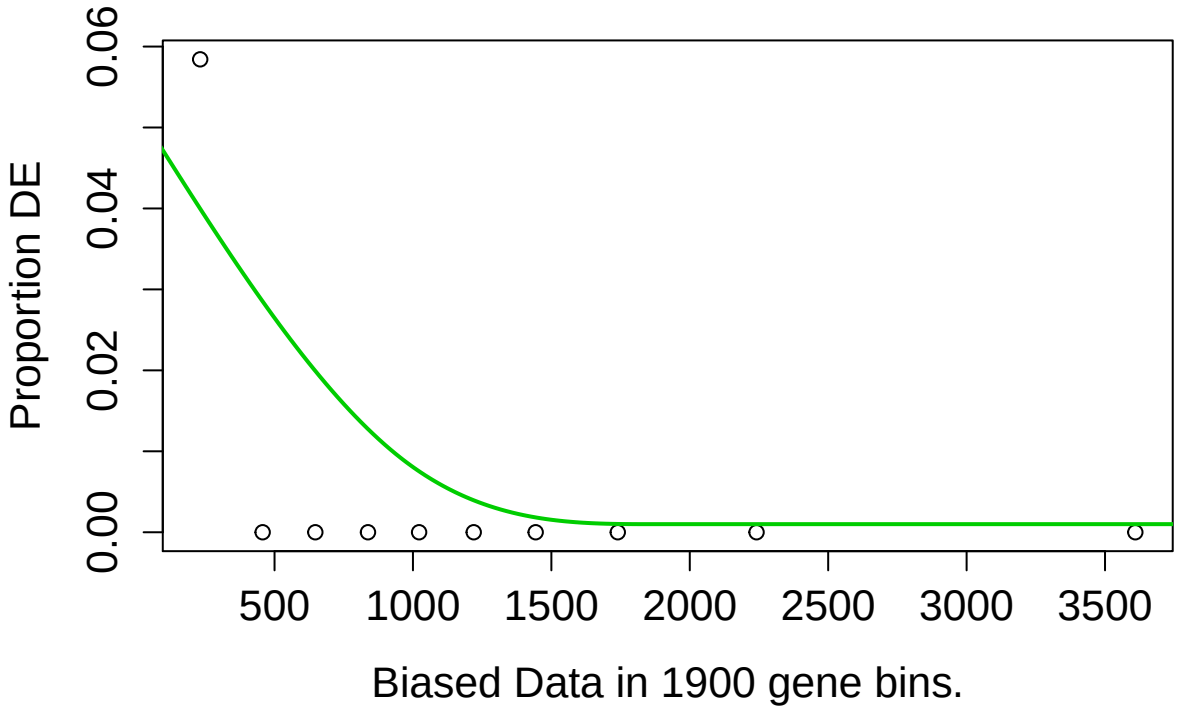



```
clusterVis_line(3)
```

```
## Using gene, curated as id variables
```

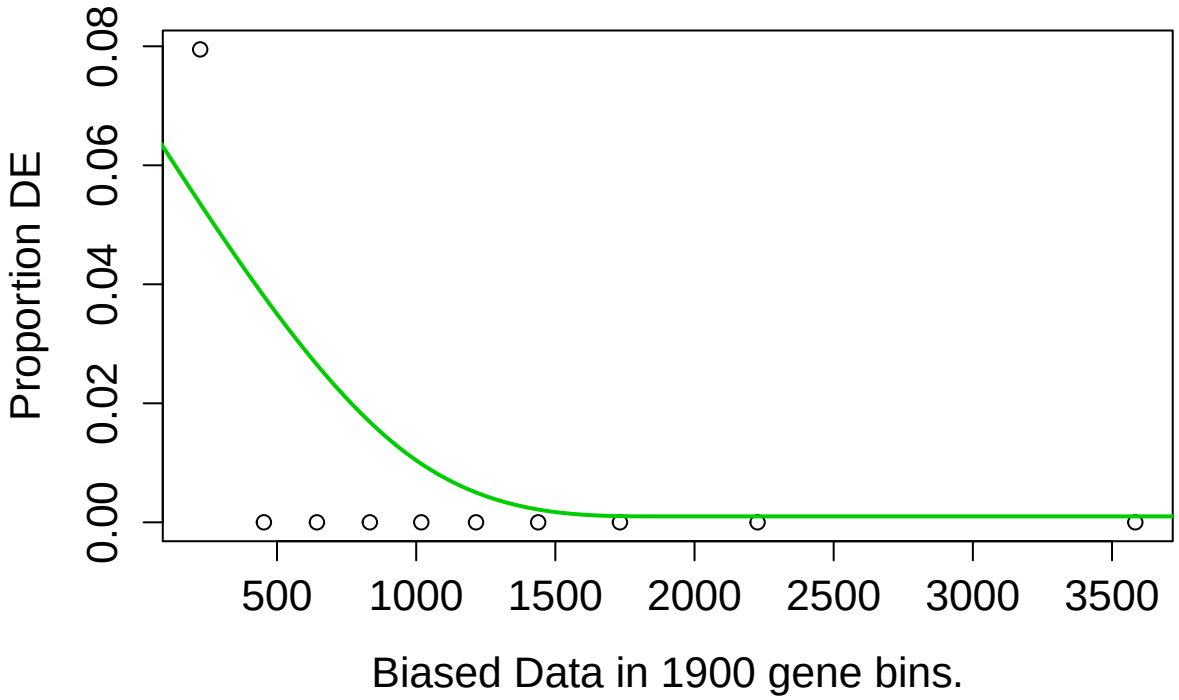



```
clusterGO(3)
```




```
## Using manually entered categories.
## For 3087 genes, we could not find any categories. These genes will be excluded.
```

```
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```

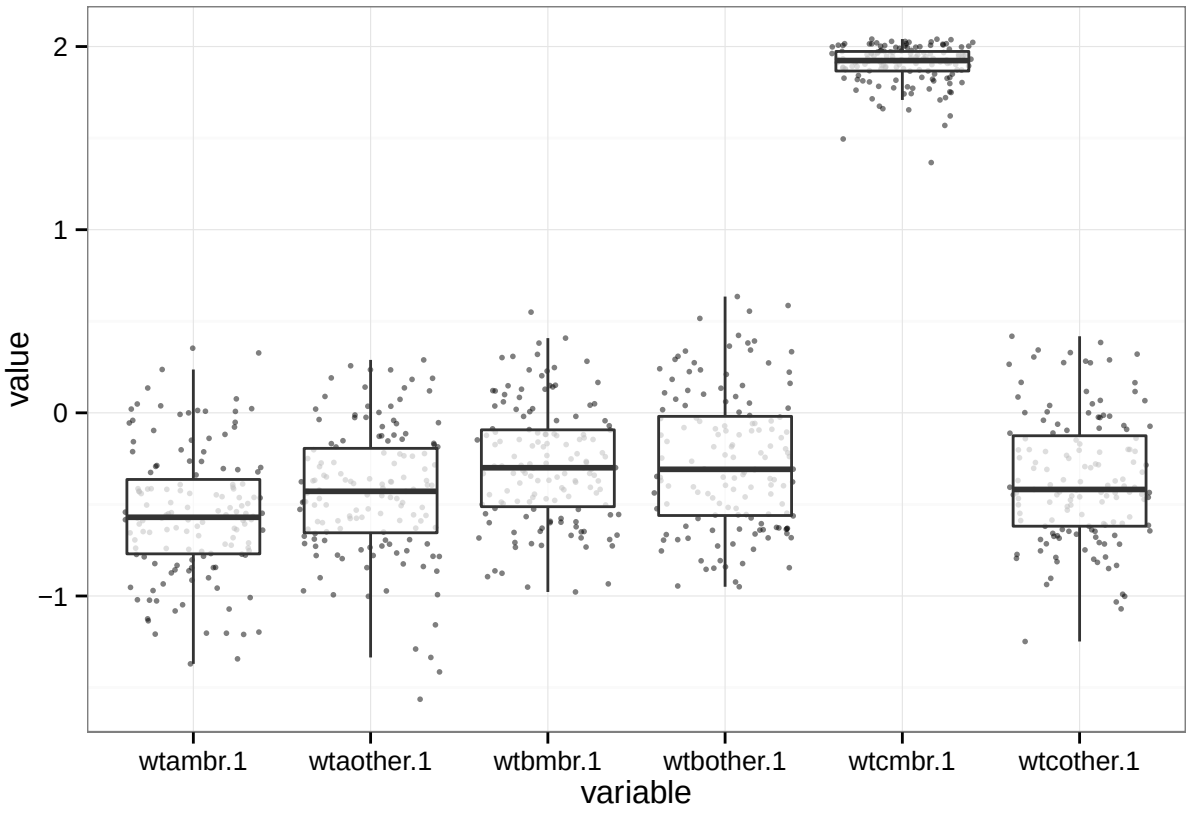



```
##      [,1]
```

**Cluster 4 - Photosynthetic Genes.**

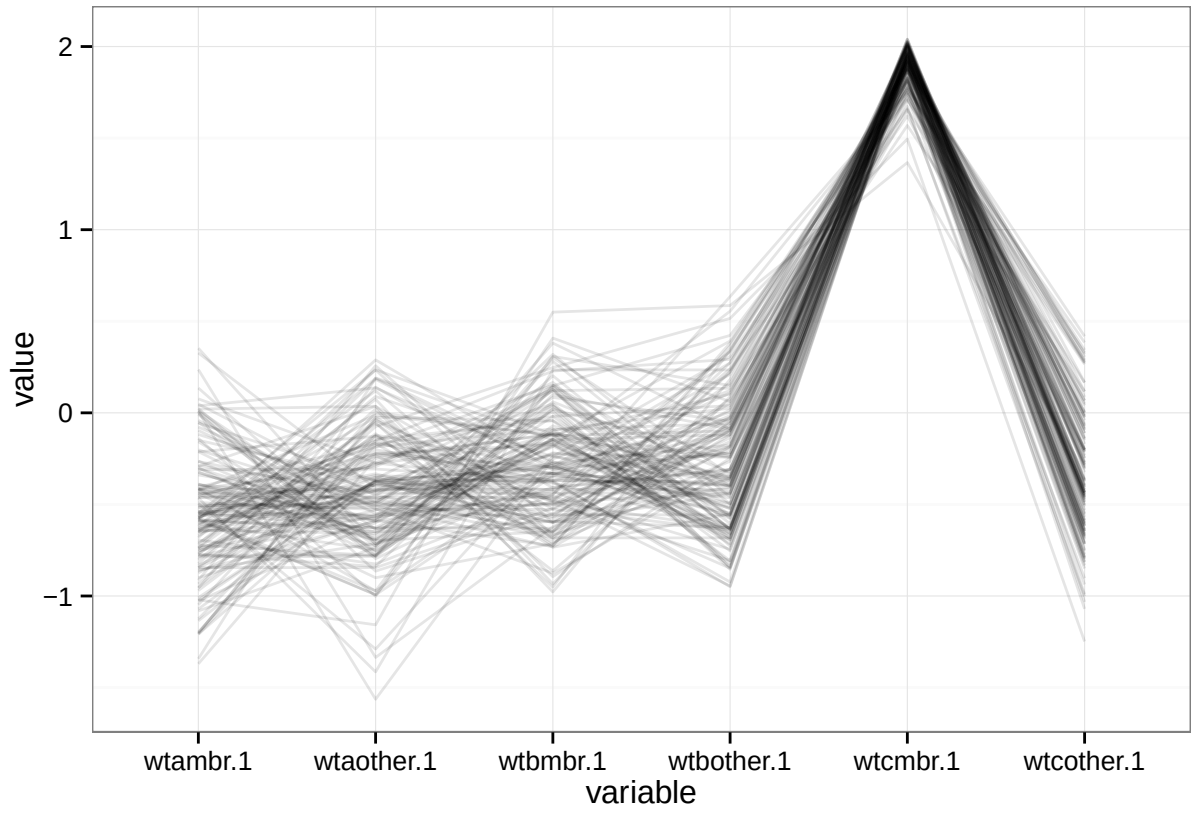
```
clusterVis(4)
```

```
## Using gene as id variables
```

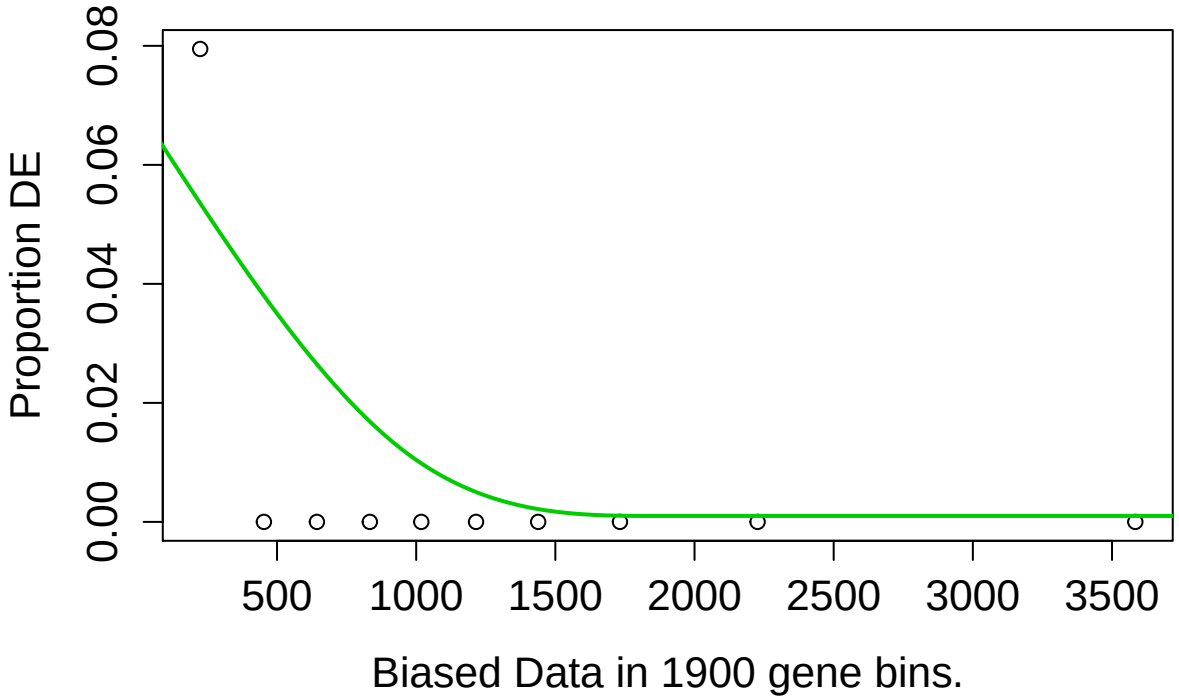



```
clusterVis_line(4)
```

```
## Using gene, curated as id variables
```

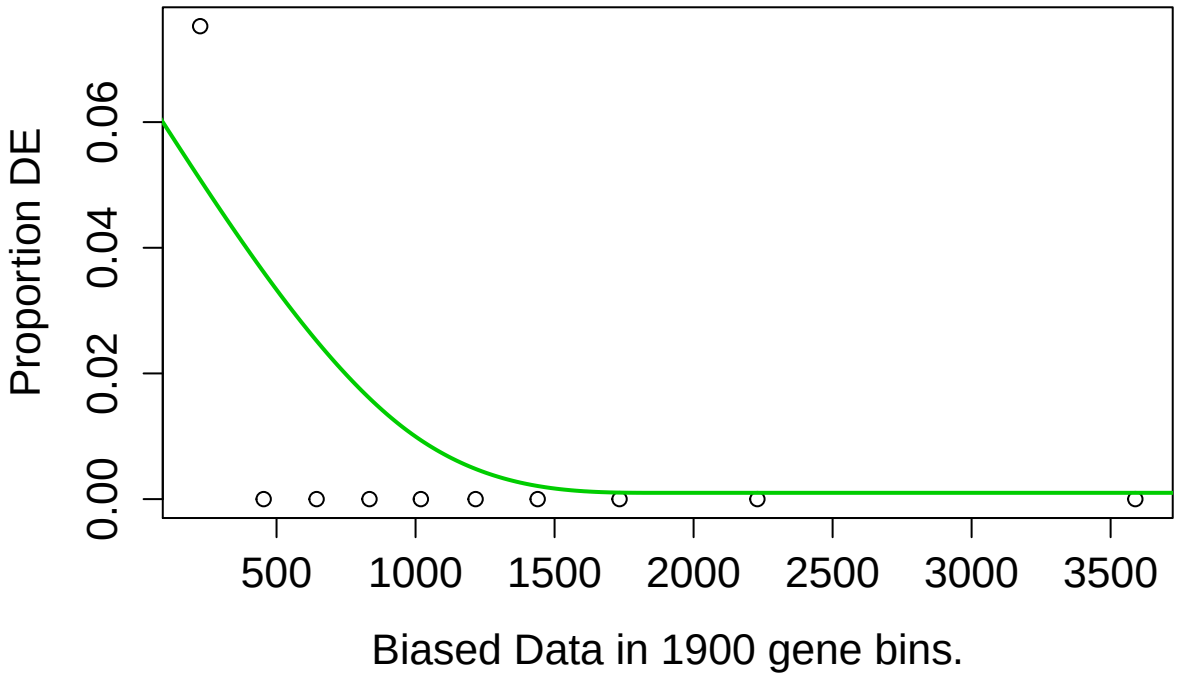



```
clusterGO(4)
```




```
## Using manually entered categories.
## For 3079 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```




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

```
y <- genesInClust(4, data.val2, annotation)
```

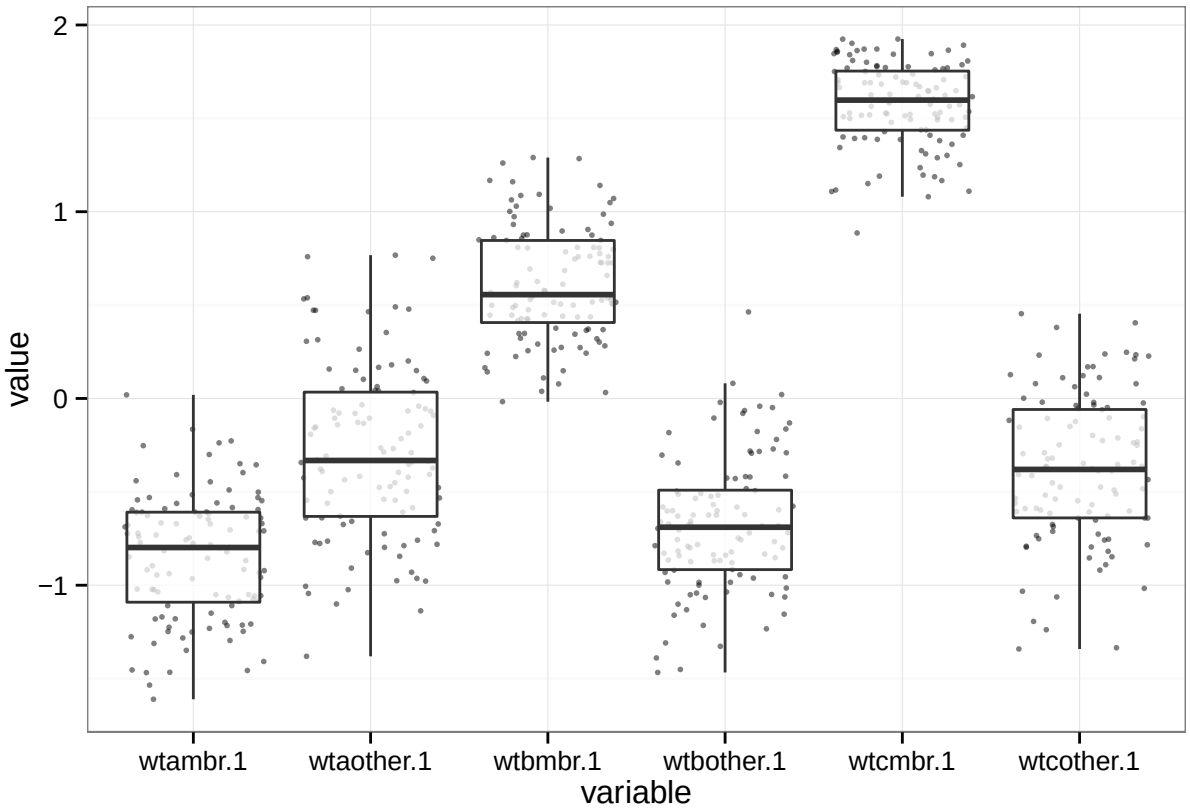
```
## [1] 125
```

**Cluster 5**

Upregulation of BOther, whith many GO categories

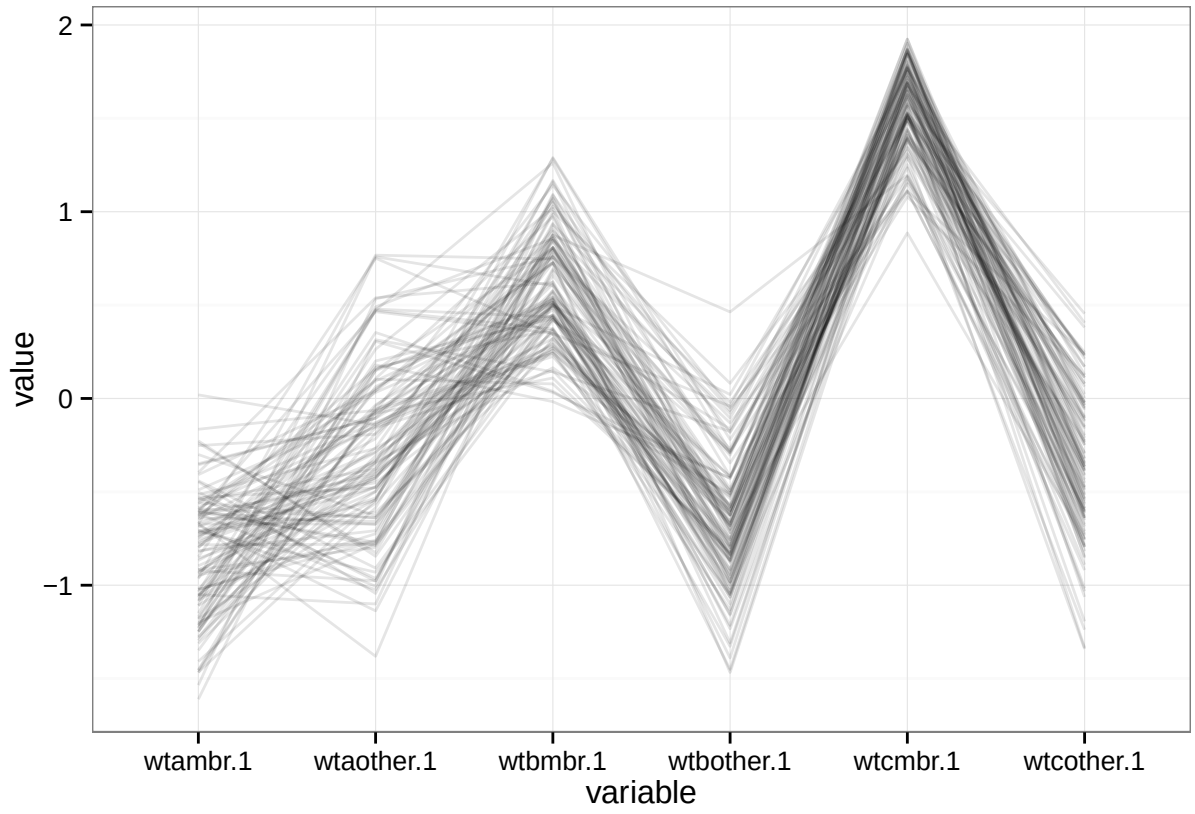
```
clusterVis(5)
```

```
## Using gene as id variables
```

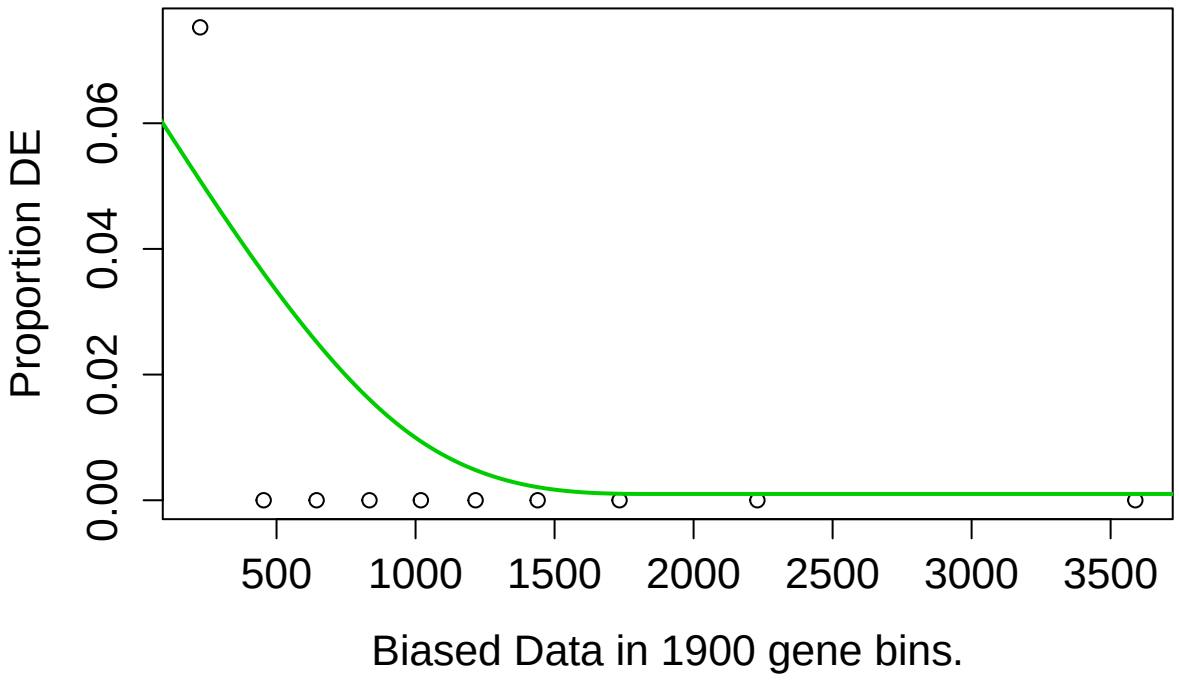



```
clusterVis_line(5)
```

```
## Using gene, curated as id variables
```

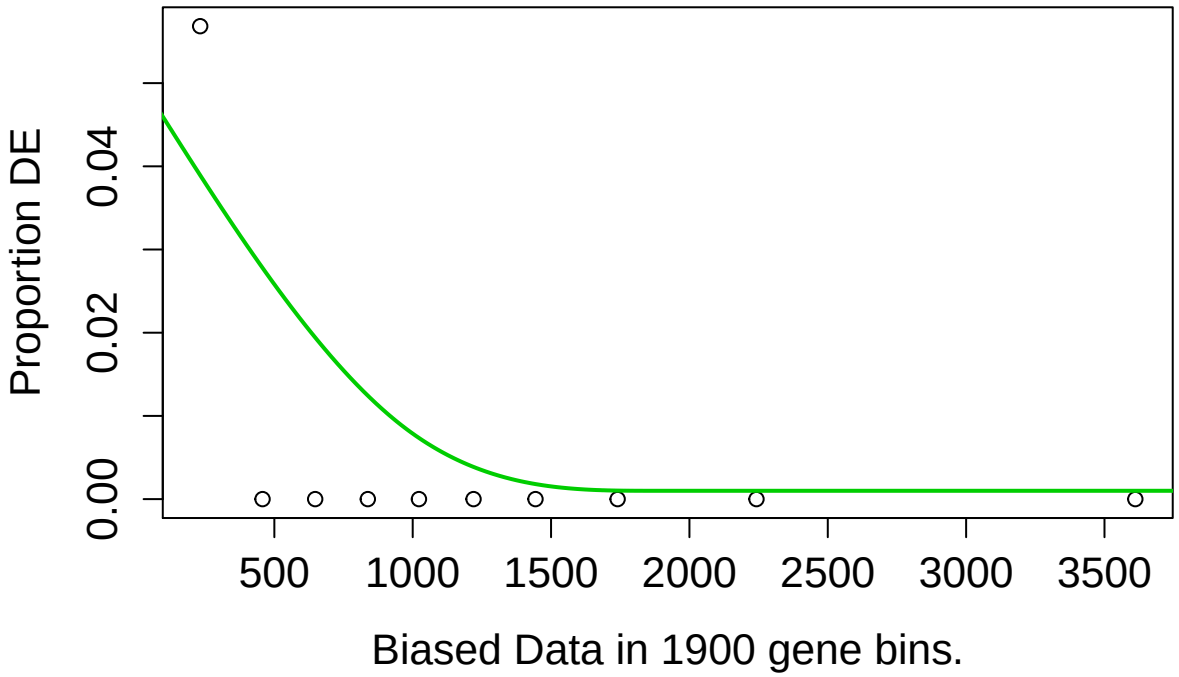



```
clusterGO(5)
```




```
## Using manually entered categories.
## For 3044 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```




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

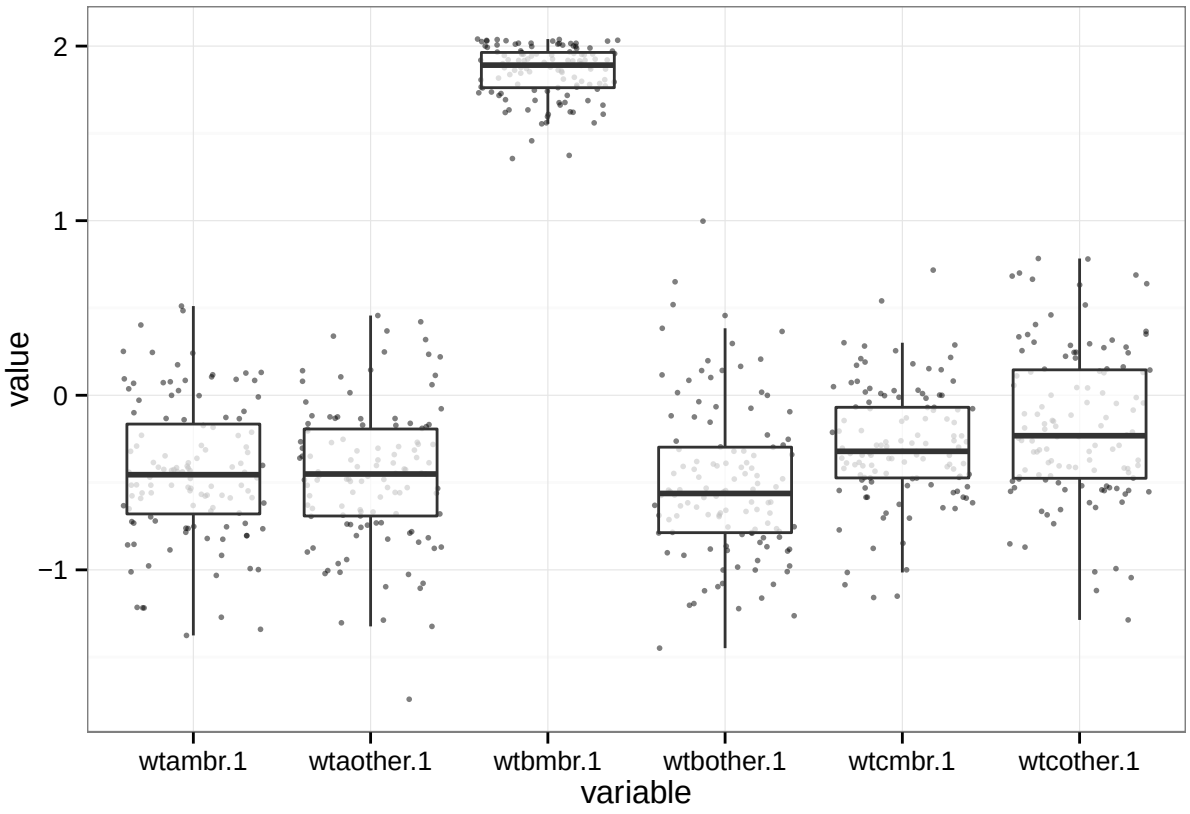
```
y <- genesInClust(5, data.val2, annotation)
```

```
## [1] 125
```

**Cluster 6**

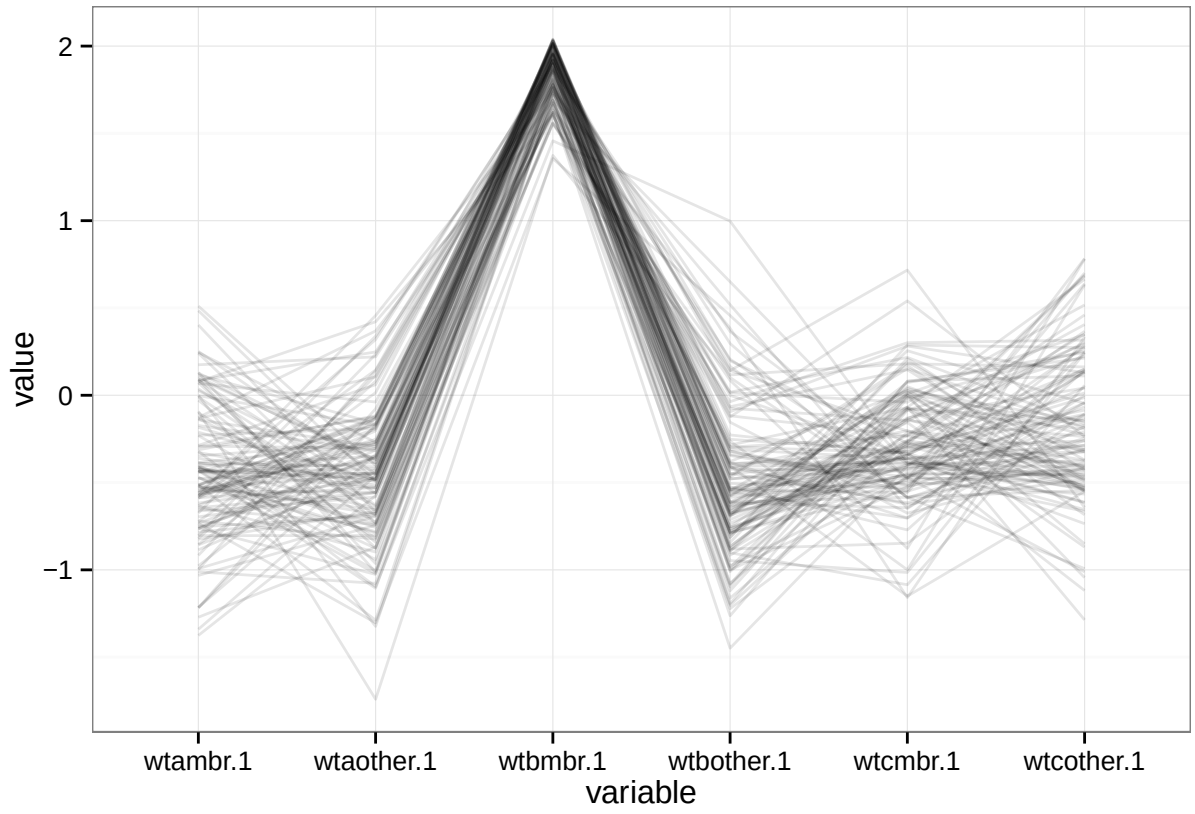
```
clusterVis(6)
```

```
## Using gene as id variables
```

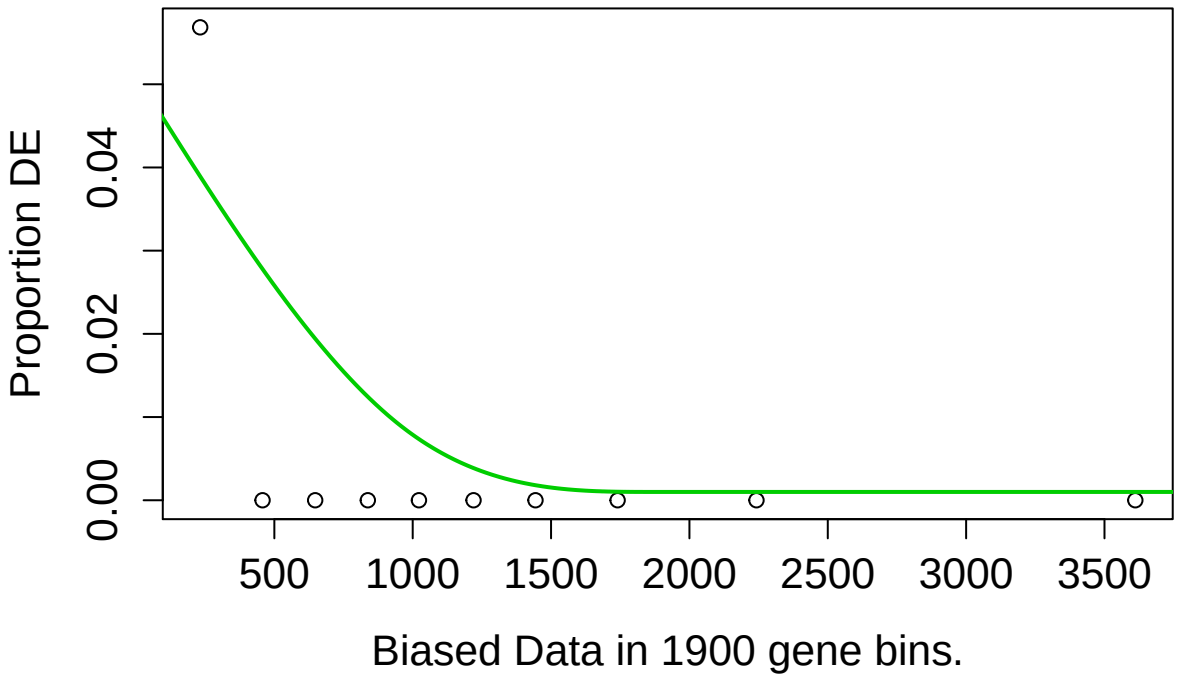



```
clusterVis_line(6)
```

```
## Using gene, curated as id variables
```

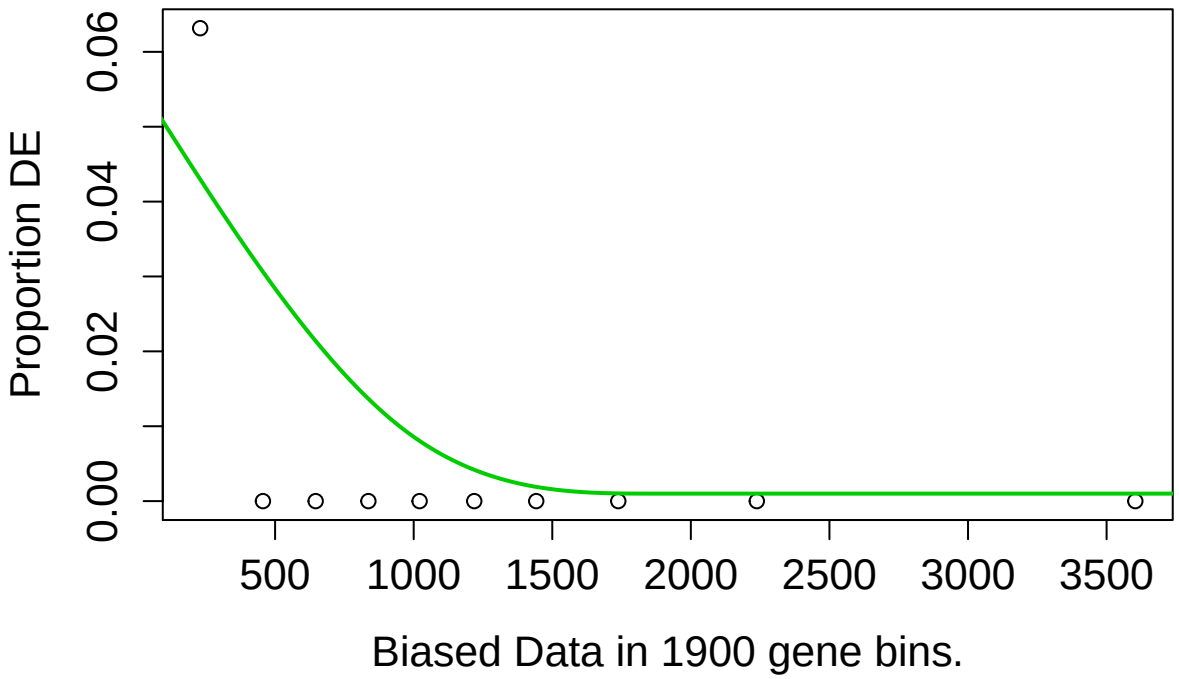



```
clusterGO(6)
```




```
## Using manually entered categories.
## For 3056 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```

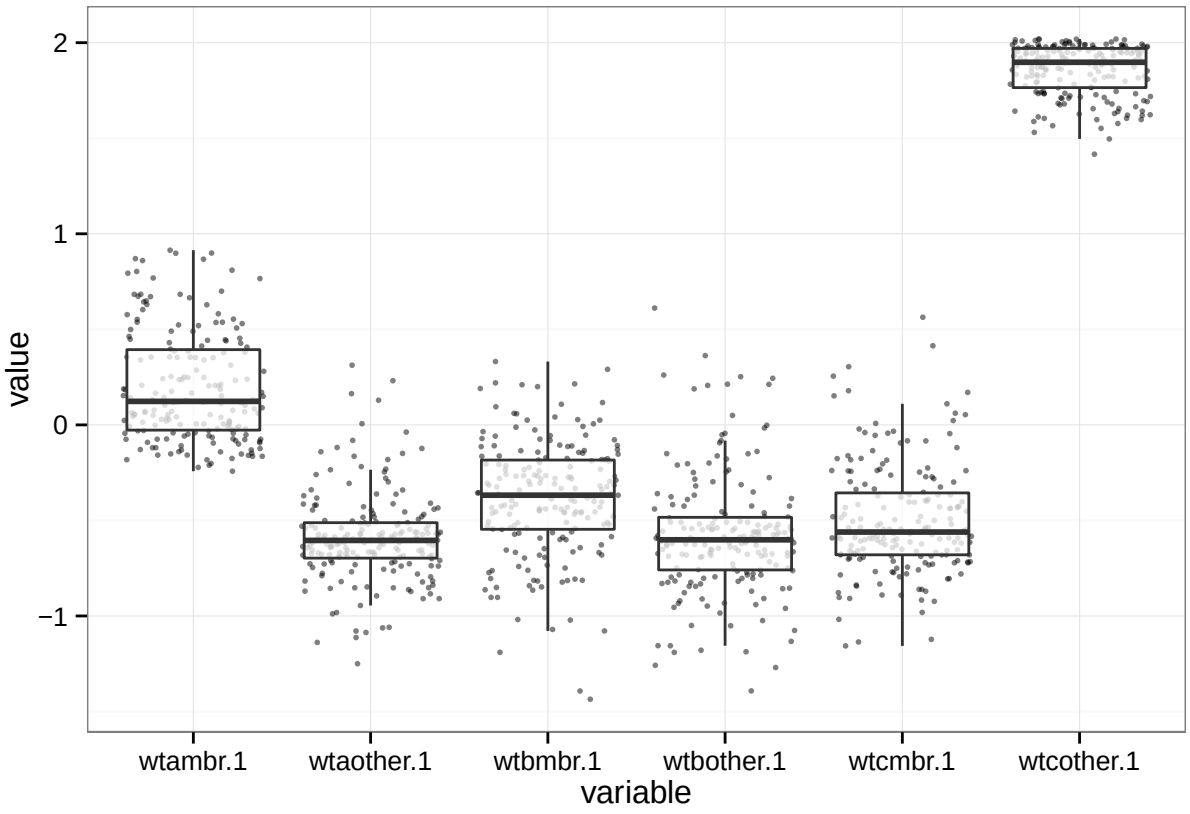



```
##      [,1]
```

**Cluster 7**

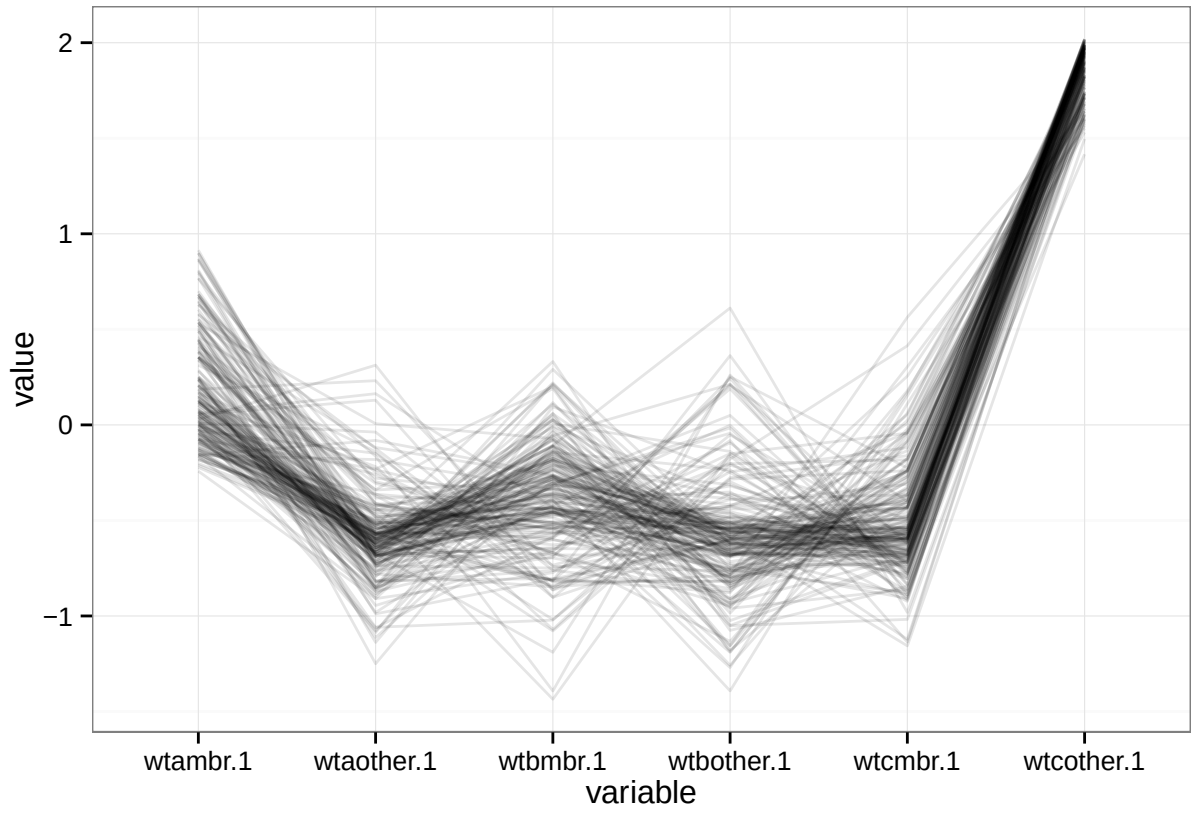
```
clusterVis(7)
```

```
## Using gene as id variables
```

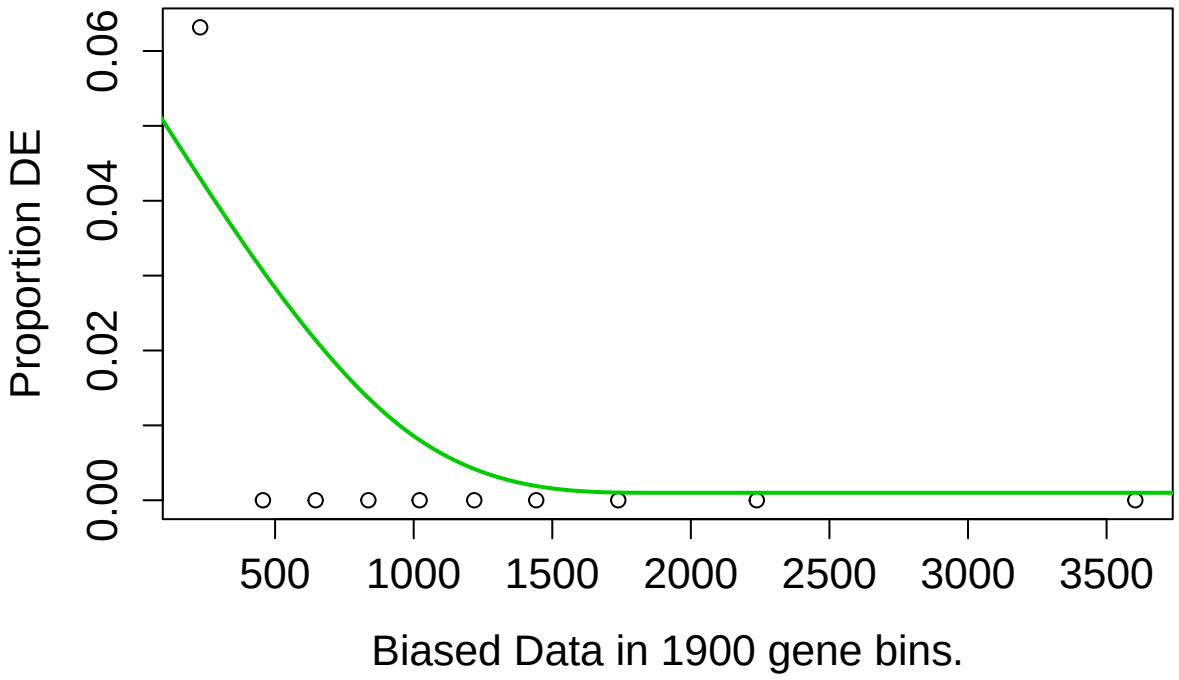



```
clusterVis_line(7)
```

```
## Using gene, curated as id variables
```

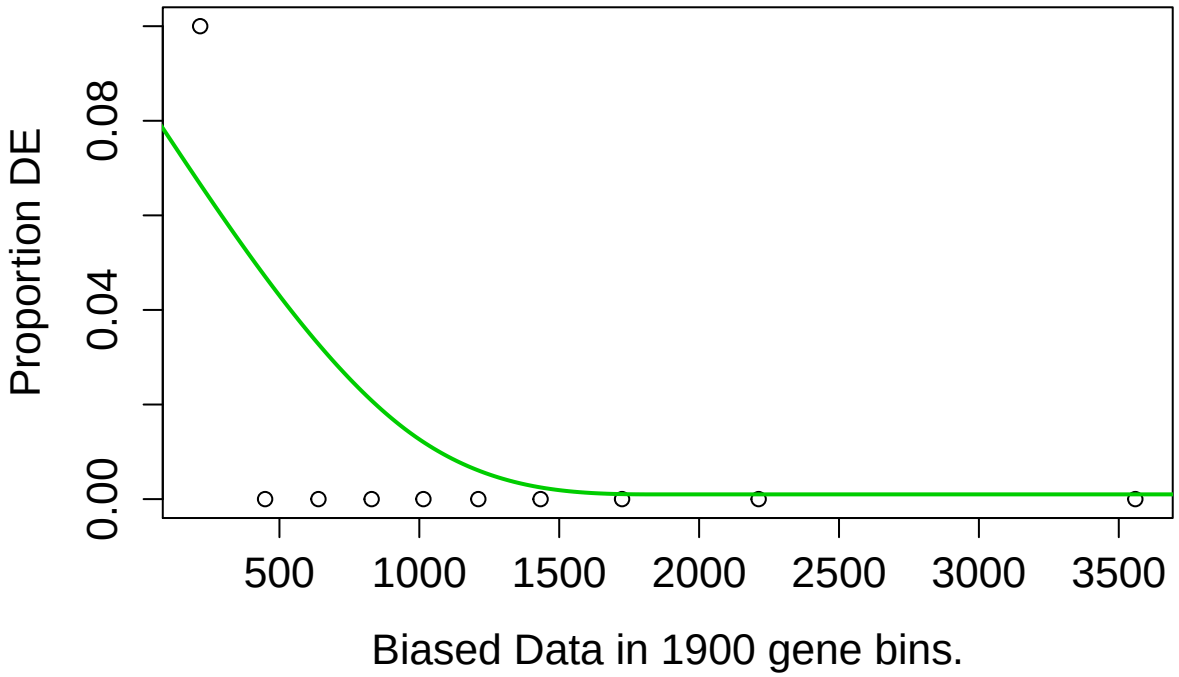



```
clusterGO(7)
```




```
## Using manually entered categories.
## For 3126 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```

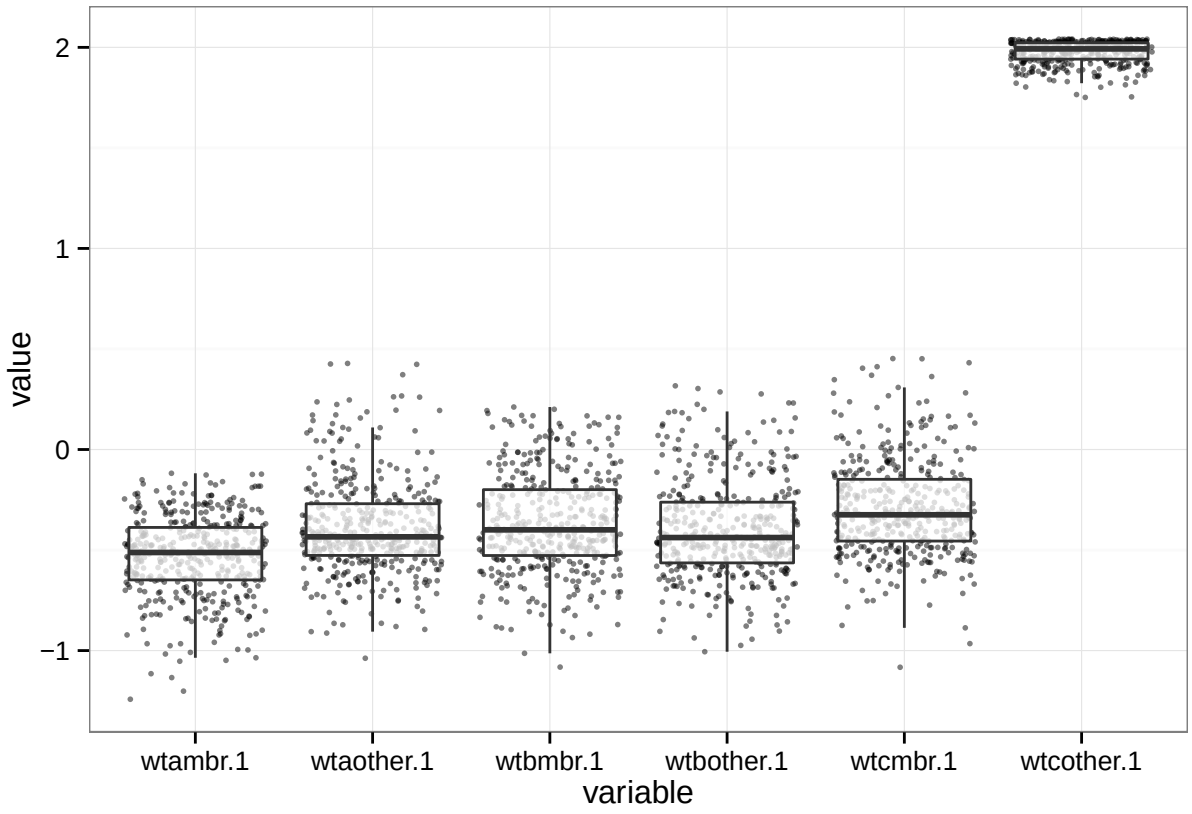



```
##      [,1]
```

**Cluster 8**

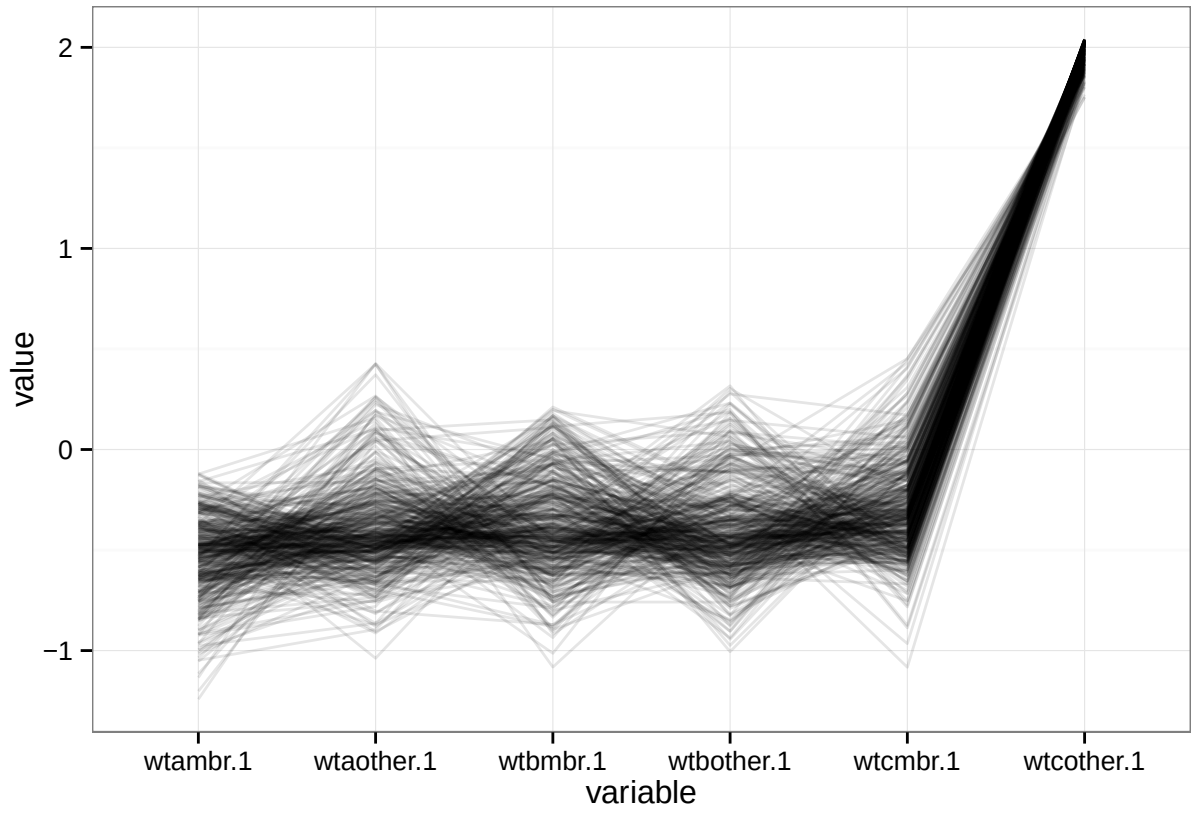
```
clusterVis(8)
```

```
## Using gene as id variables
```

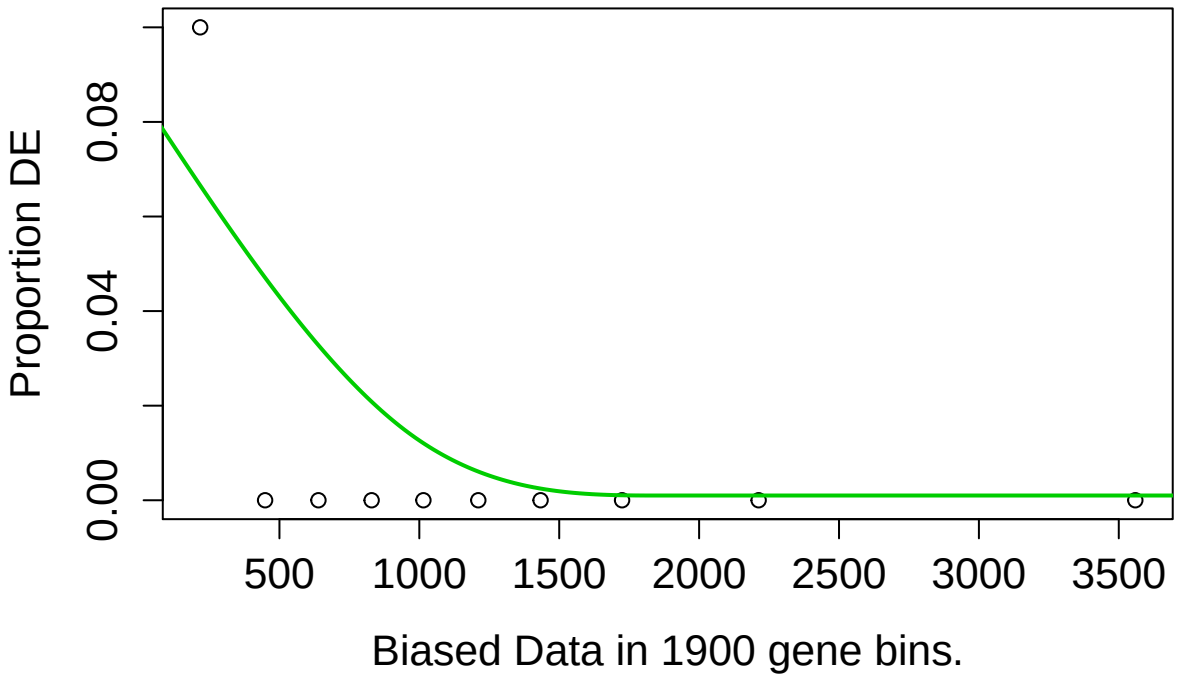



```
clusterVis_line(8)
```

```
## Using gene, curated as id variables
```

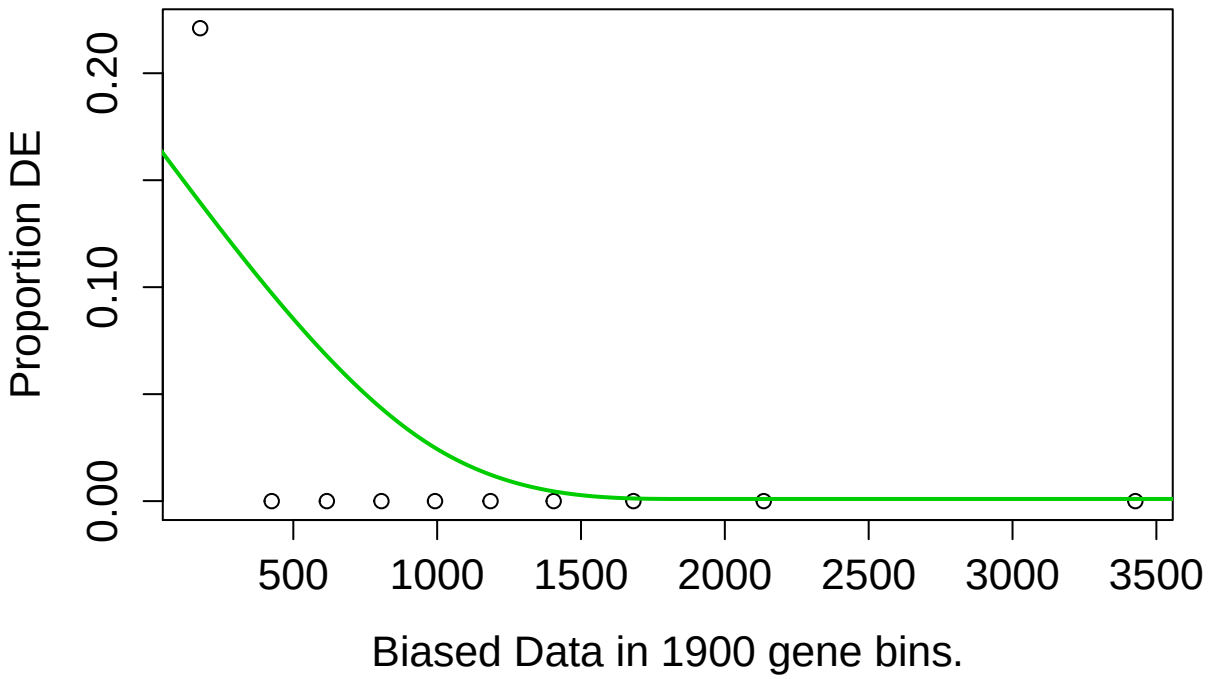



```
clusterGO(8)
```




```
## Using manually entered categories.
## For 3356 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```

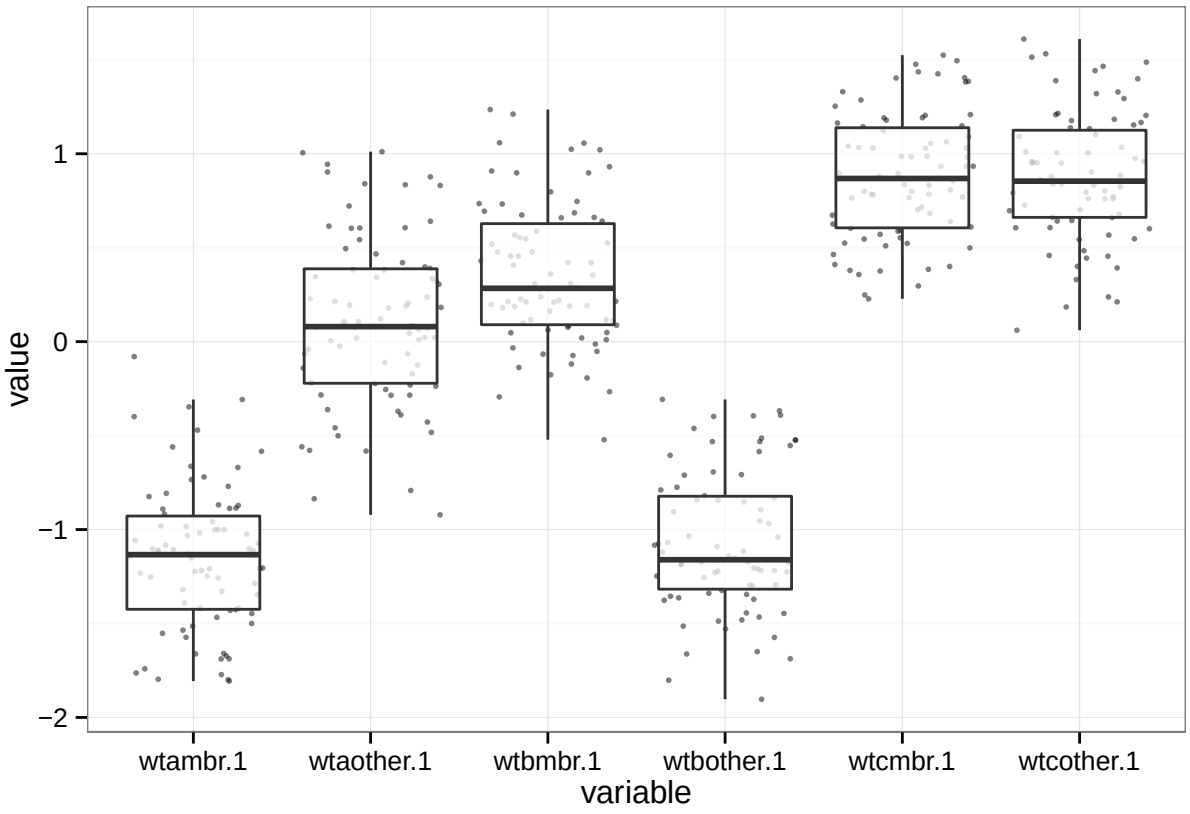



```
##      [,1]
```

**Cluster 9**

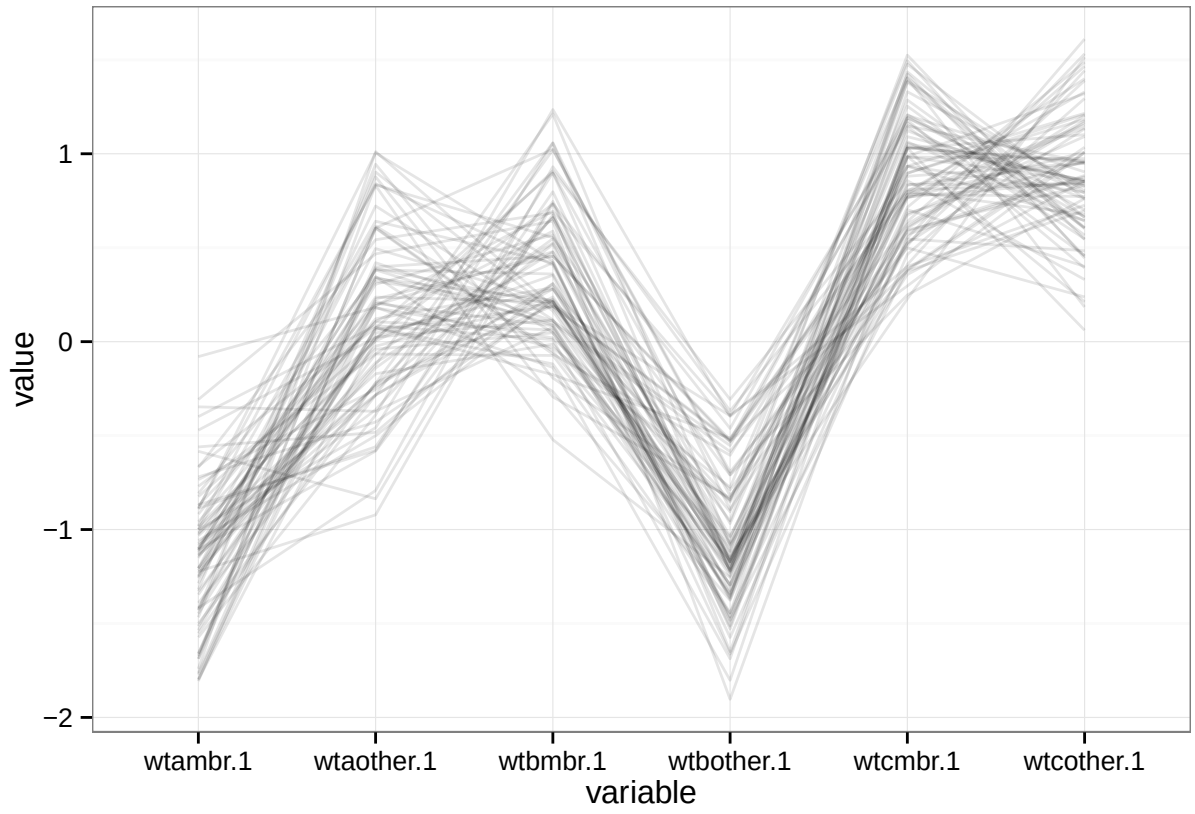
```
clusterVis(9)
```

```
## Using gene as id variables
```

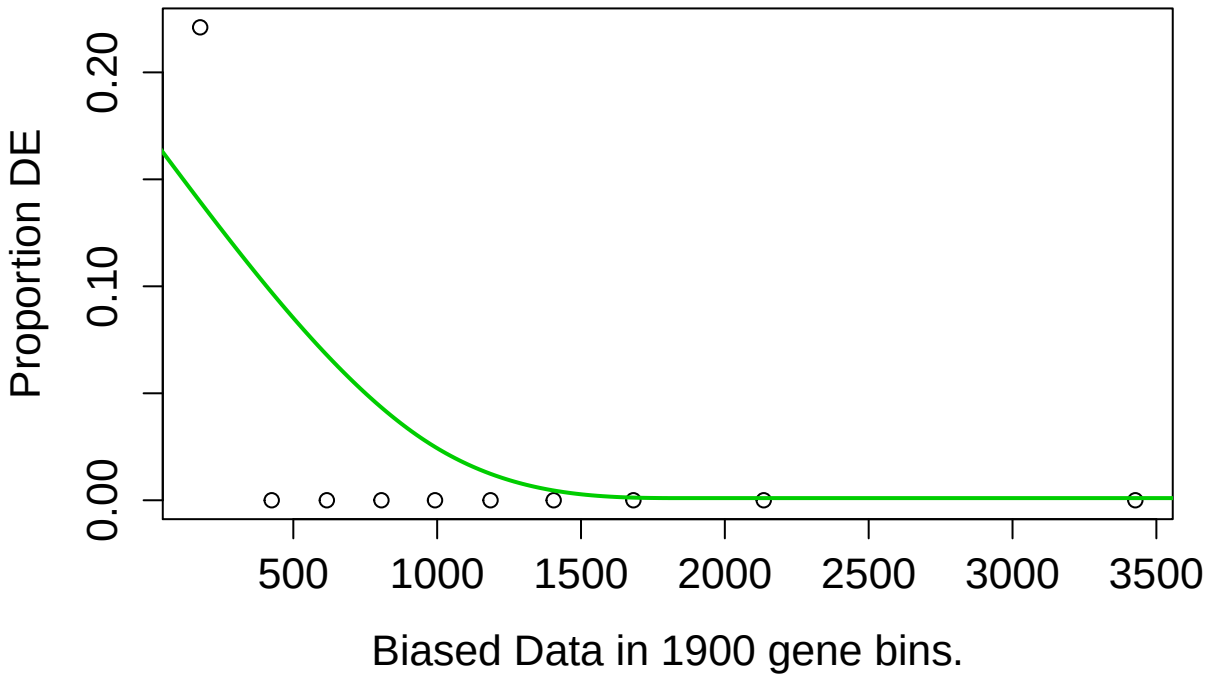



```
clusterVis_line(9)
```

```
## Using gene, curated as id variables
```

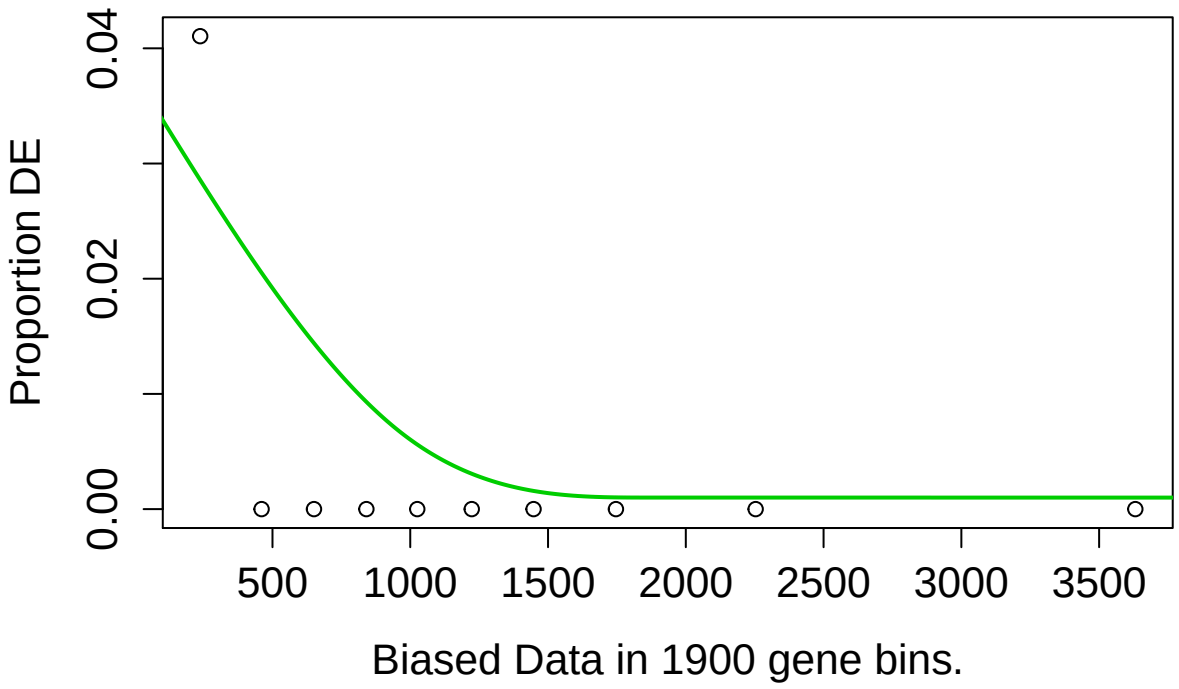



```
clusterGO(9)
```




```
## Using manually entered categories.
## For 3014 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```

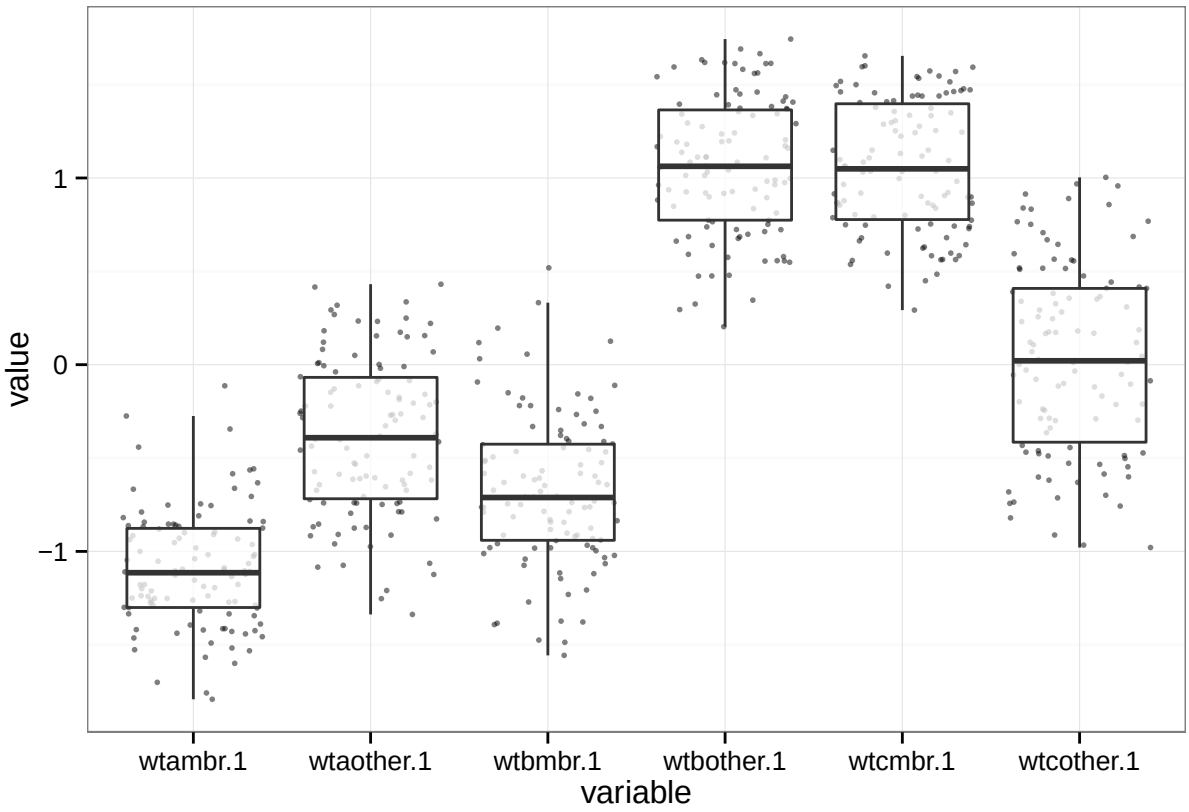



```
##      [,1]
```

**Cluster 10**

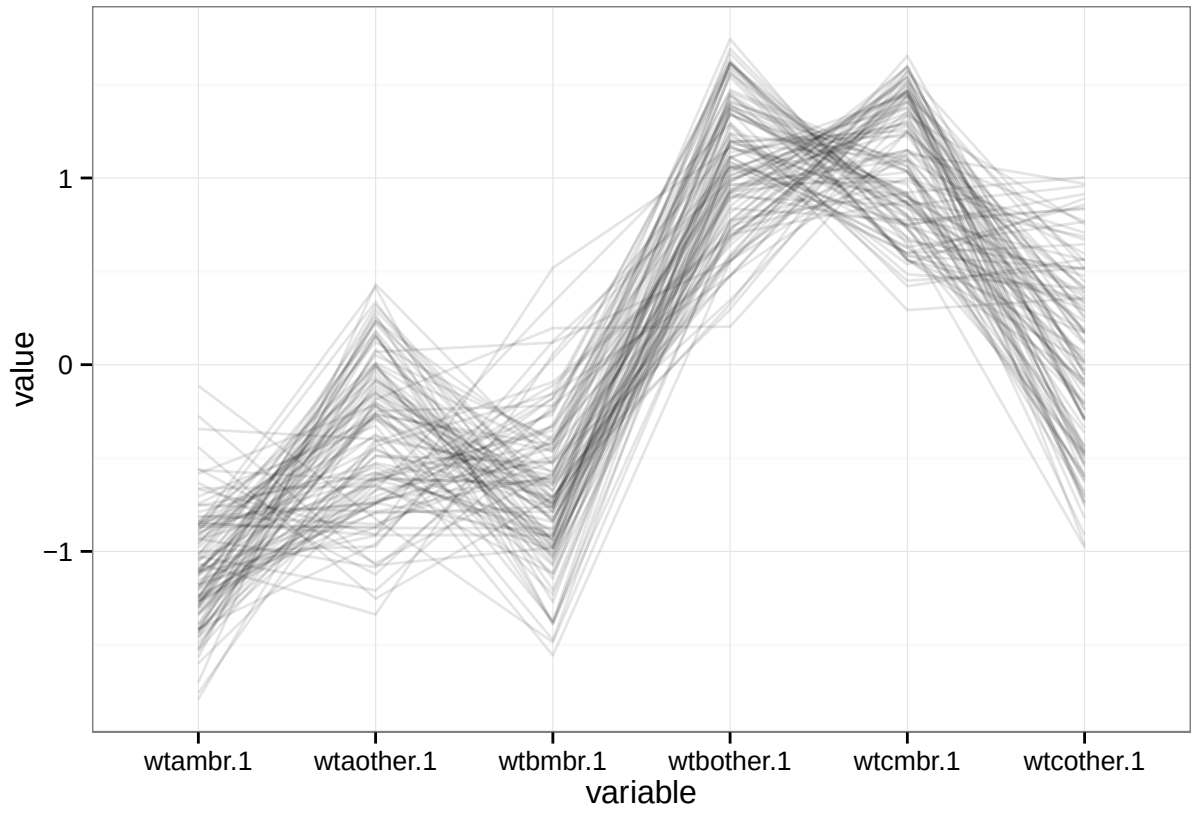
```
clusterVis(10)
```

```
## Using gene as id variables
```

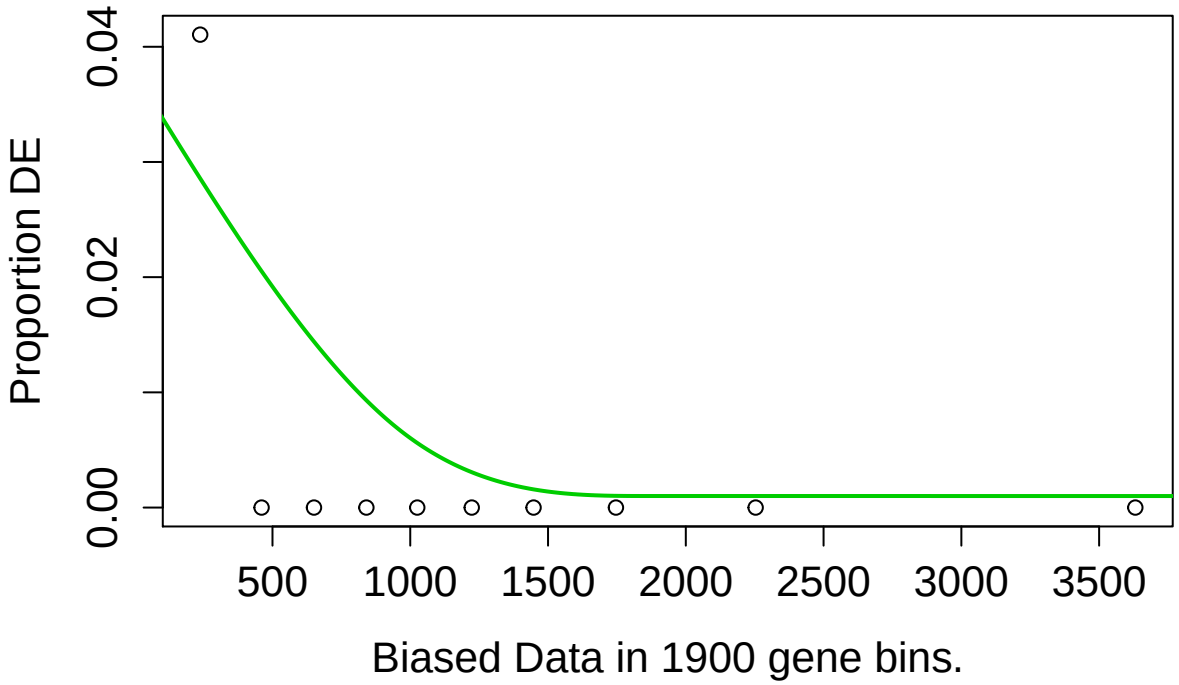



```
clusterVis_line(10)
```

```
## Using gene, curated as id variables
```

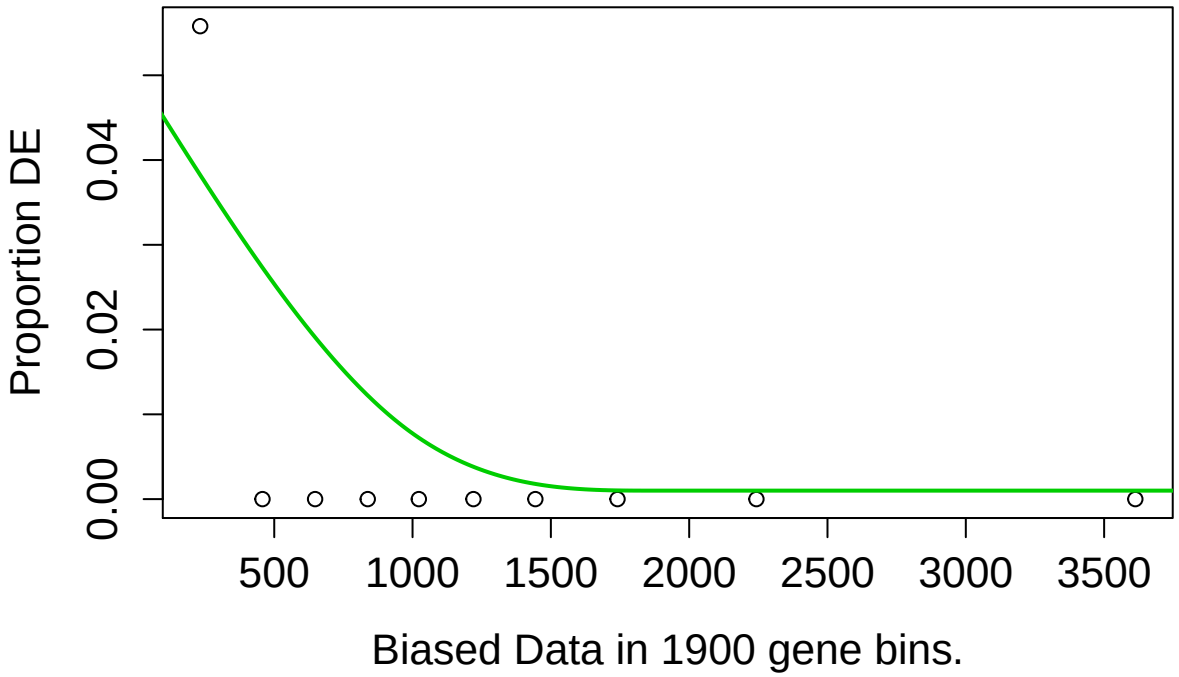



```
clusterGO(10)
```




```
## Using manually entered categories.
## For 3042 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```




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

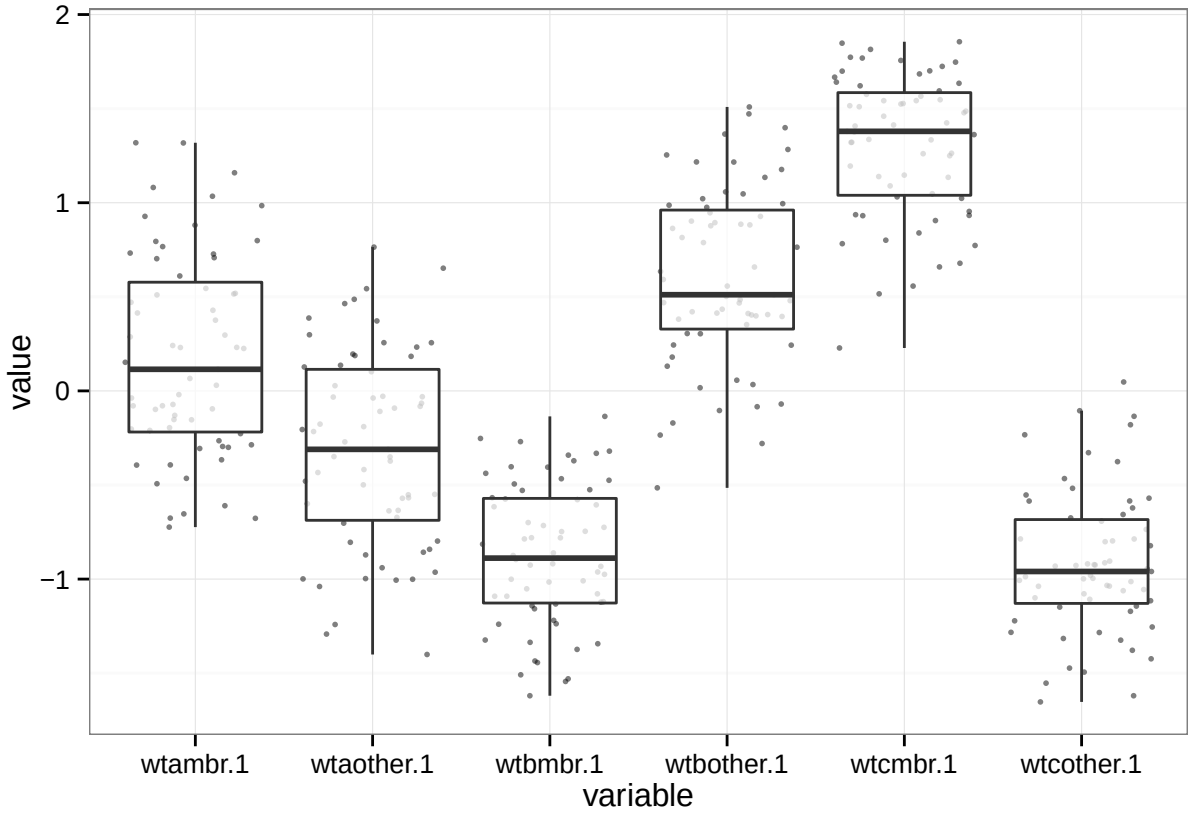
```
y <- genesInClust(10, data.val2, annotation)
```

```
## [1] 125
```

**Cluster 11**

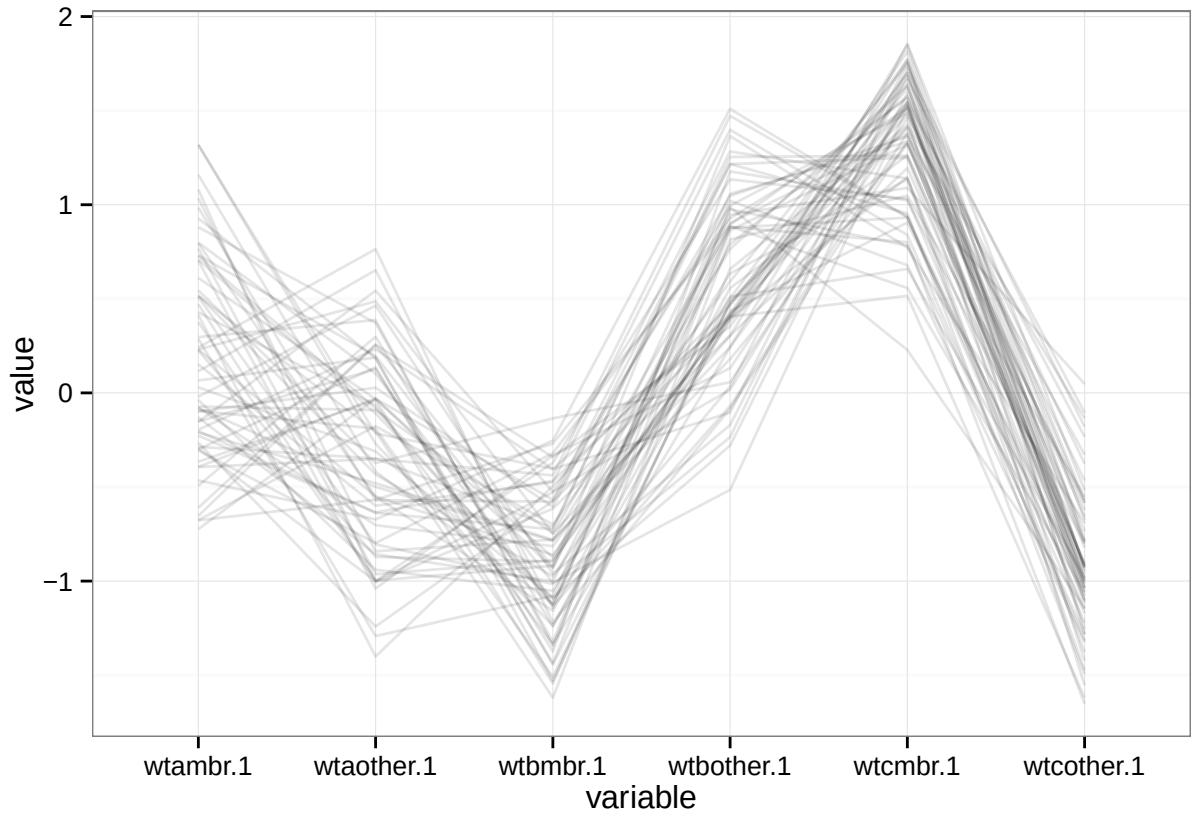
```
clusterVis(11)
```

```
## Using gene as id variables
```

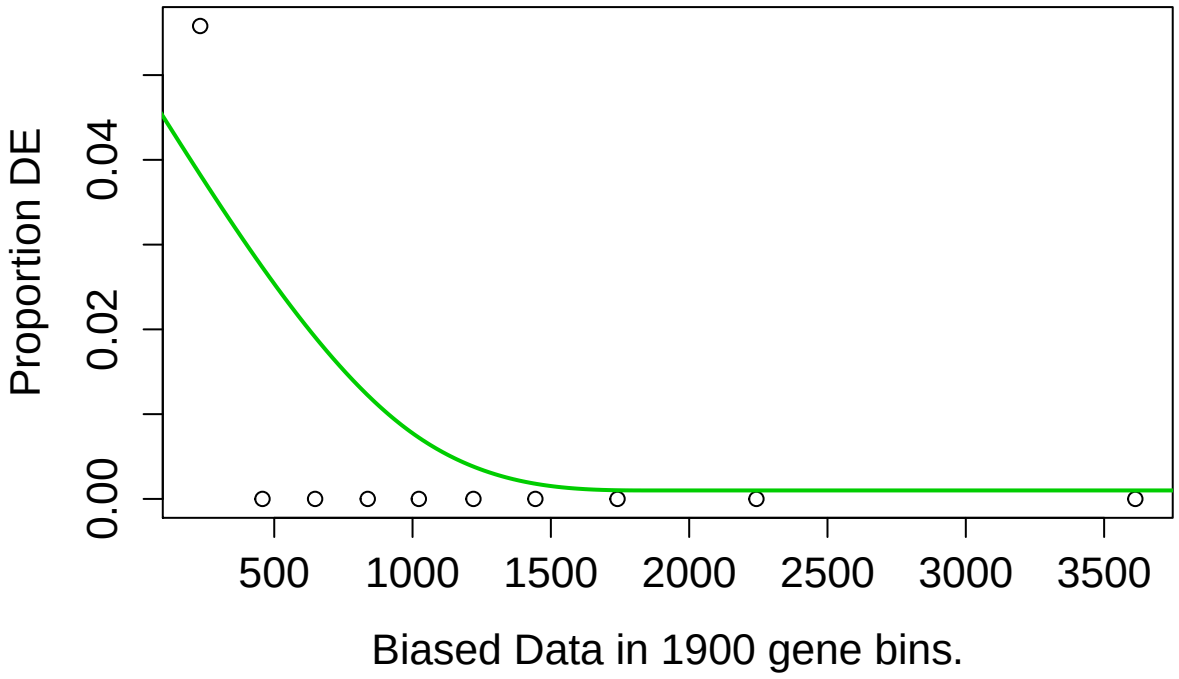



```
clusterVis_line(11)
```

```
## Using gene, curated as id variables
```

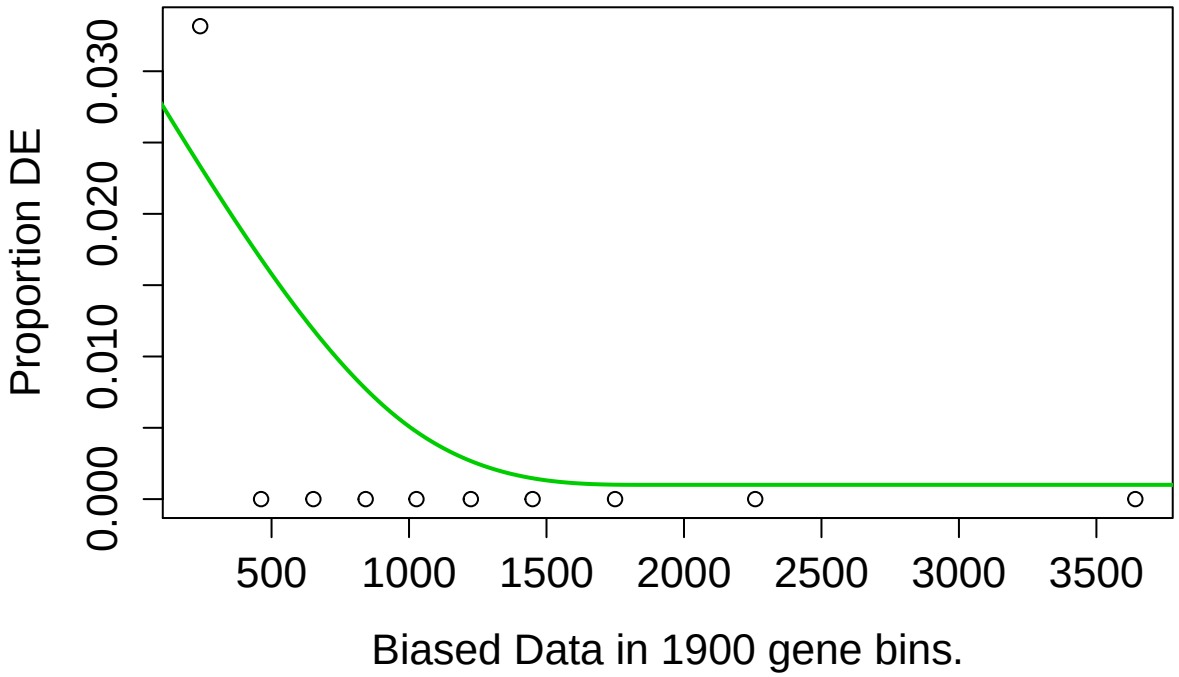



```
clusterGO(11)
```




```
## Using manually entered categories.
## For 2999 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```



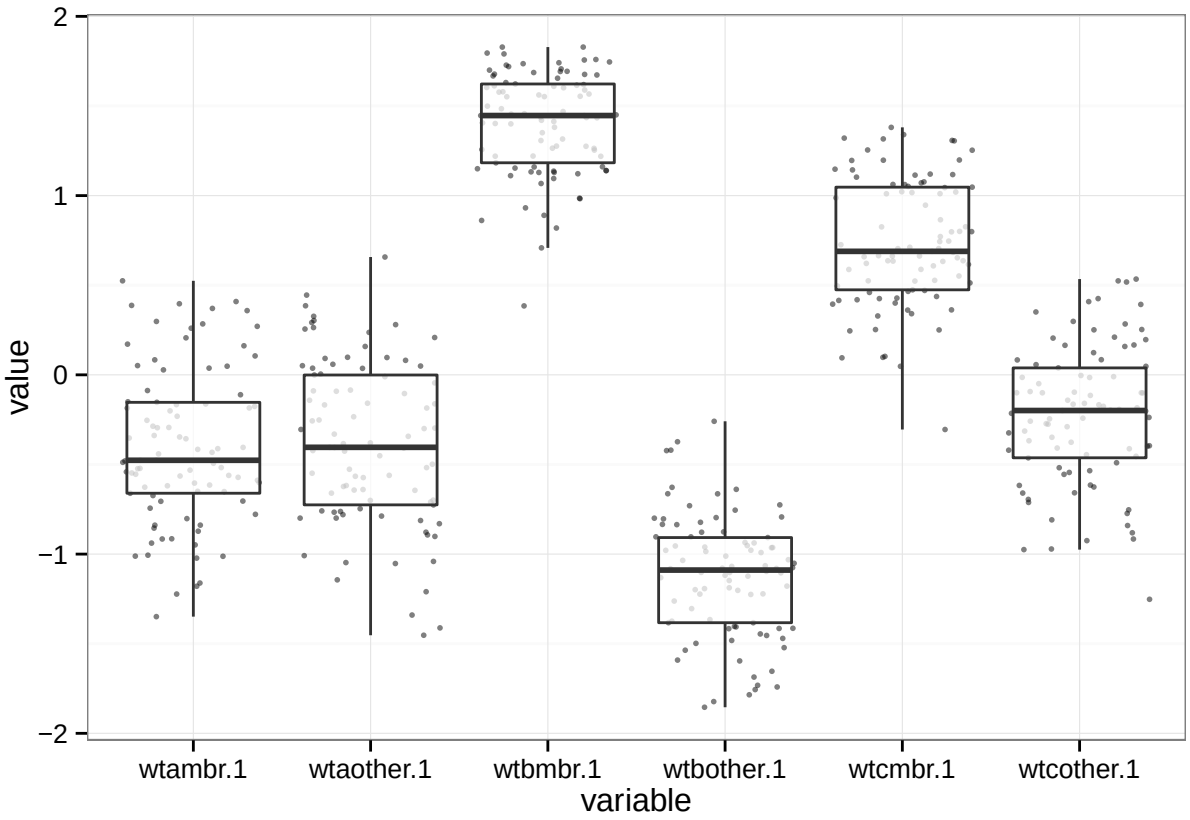

```
##      [,1]
```

```
y <- genesInClust(11, data.val2, annotation)
```

```
## [1] 125
```

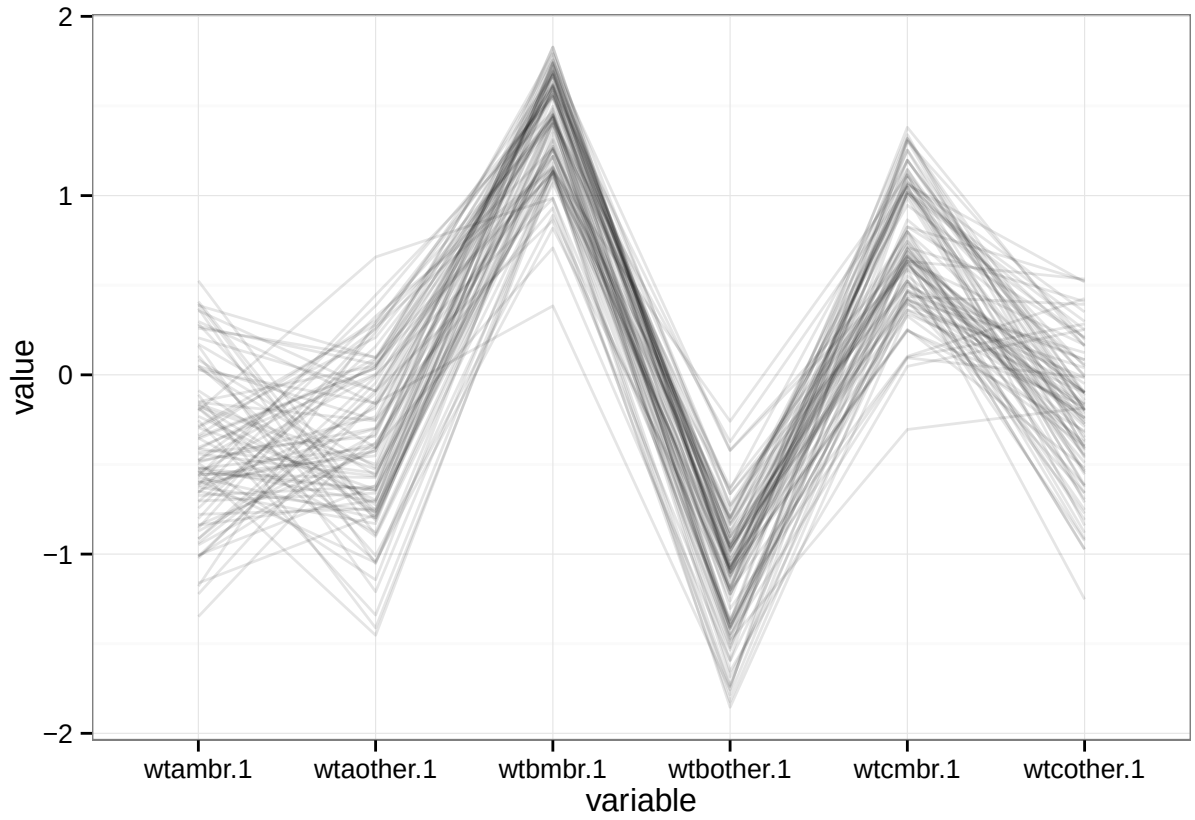
```
clusterVis(12)
```

```
## Using gene as id variables
```

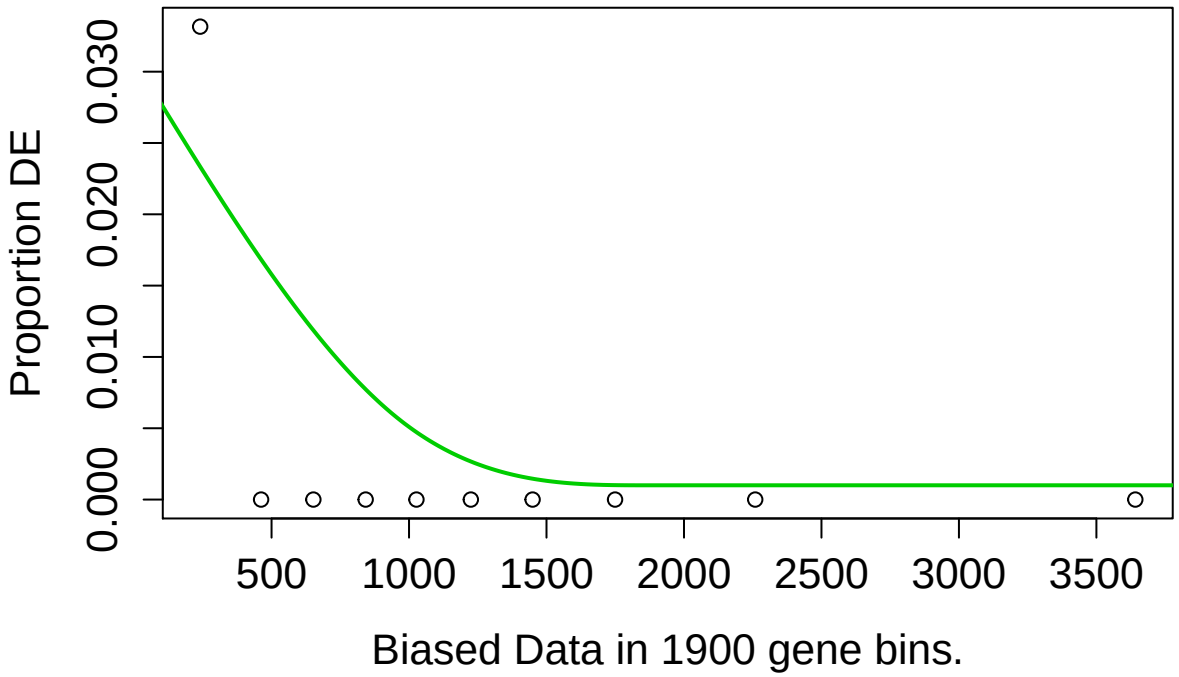



```
clusterVis_line(12)
```

```
## Using gene, curated as id variables
```

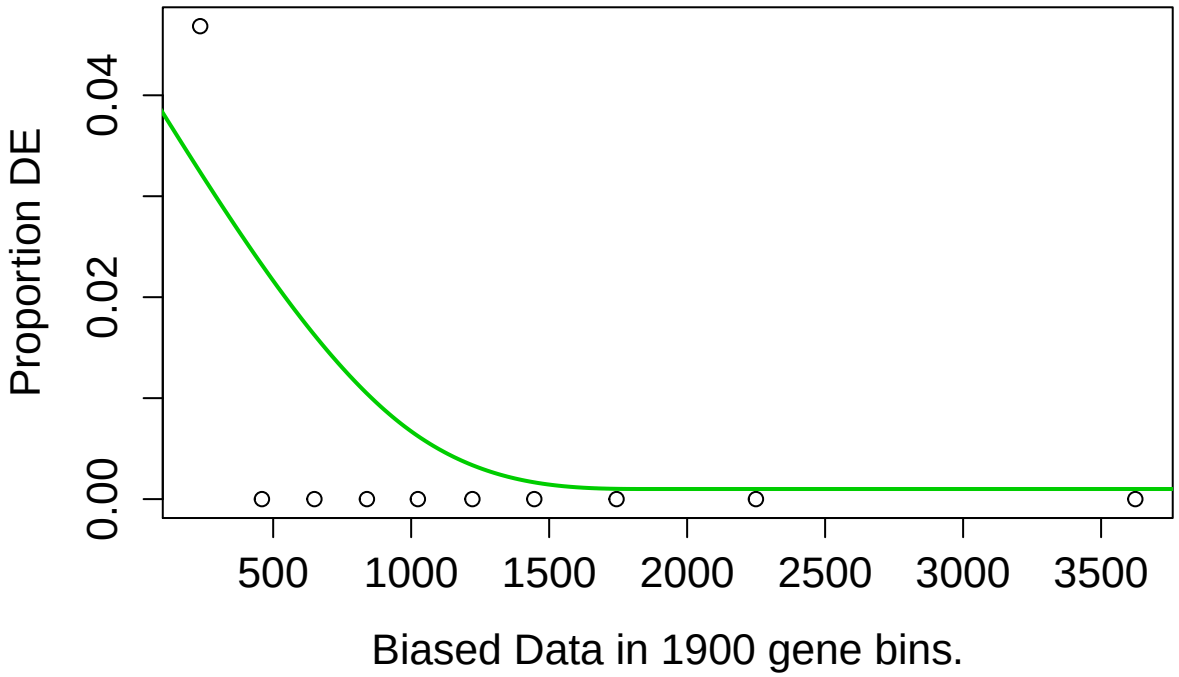



```
clusterGO(12)
```




```
## Using manually entered categories.
## For 3025 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```



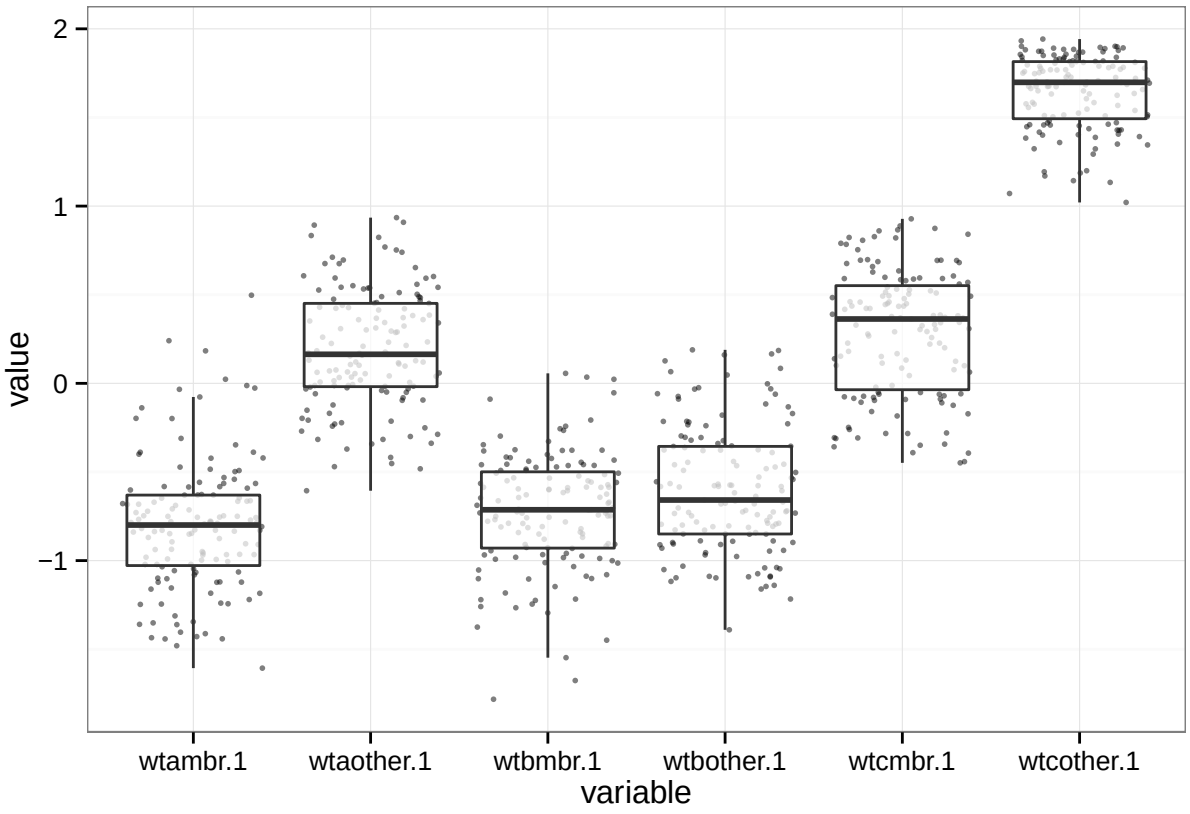

```
##      [,1]
```

```
y <- genesInClust(12, data.val2, annotation)
```

```
## [1] 125
```

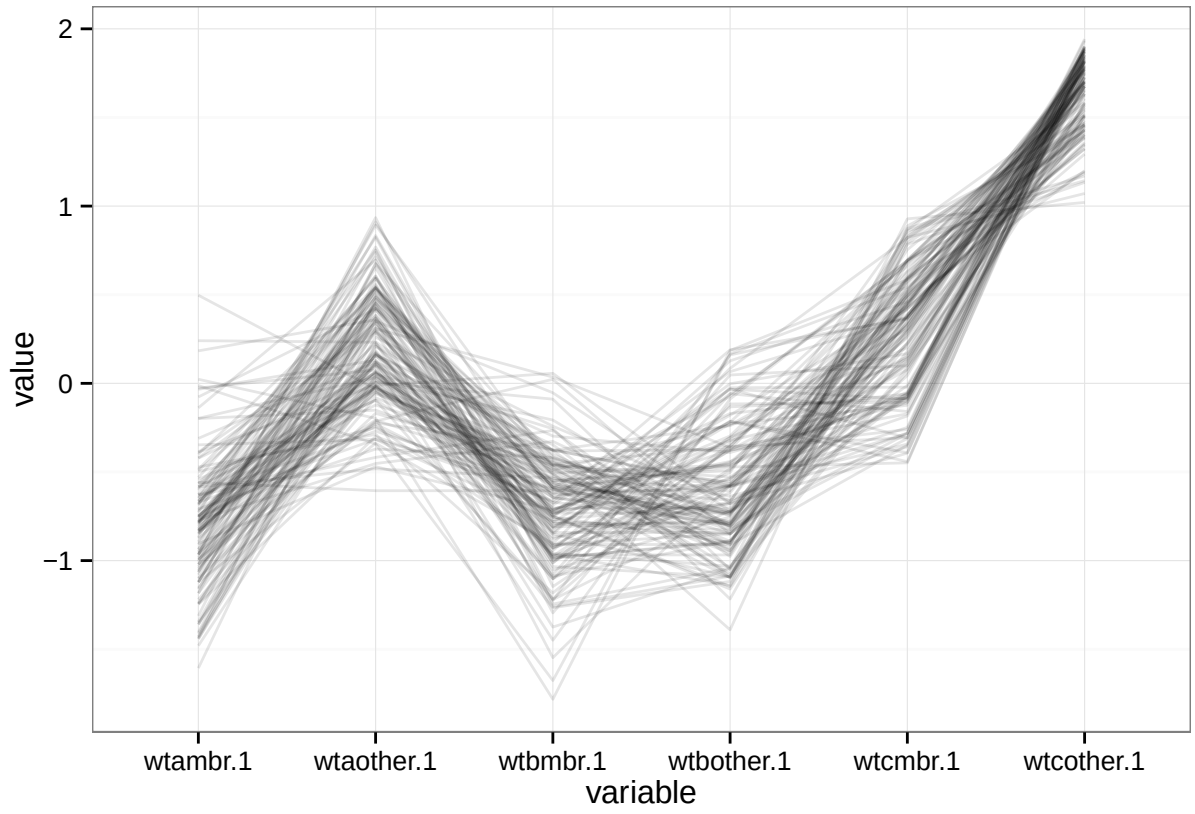
```
clusterVis(13)
```

```
## Using gene as id variables
```

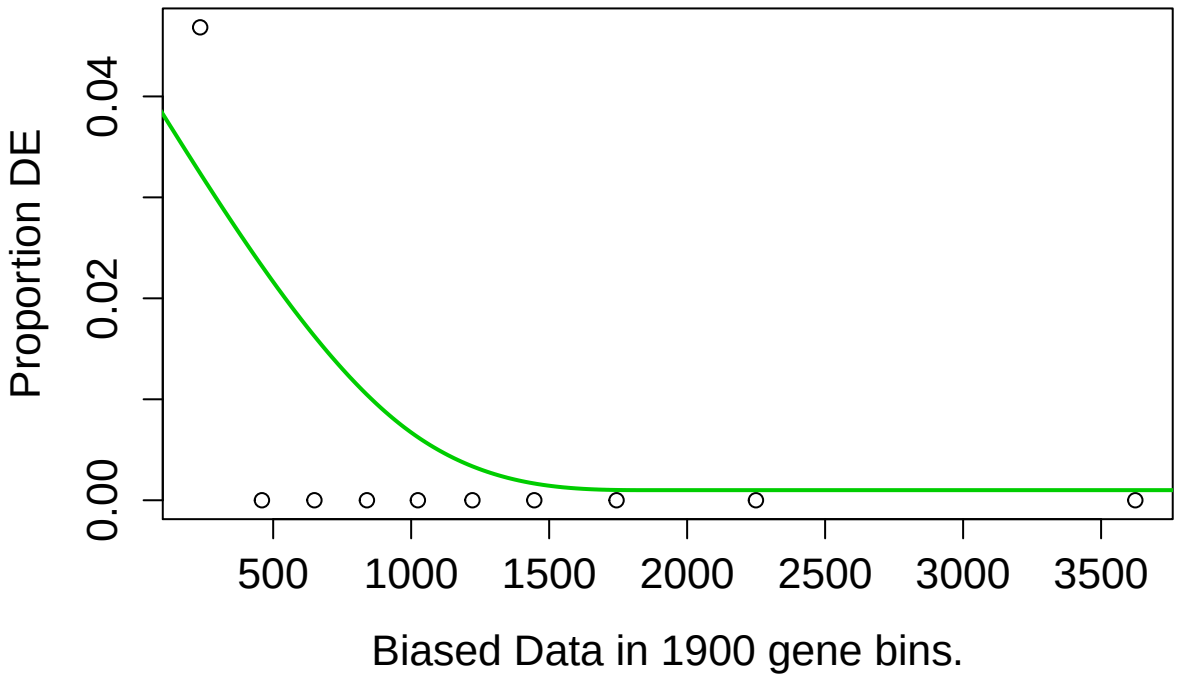



```
clusterVis_line(13)
```

```
## Using gene, curated as id variables
```

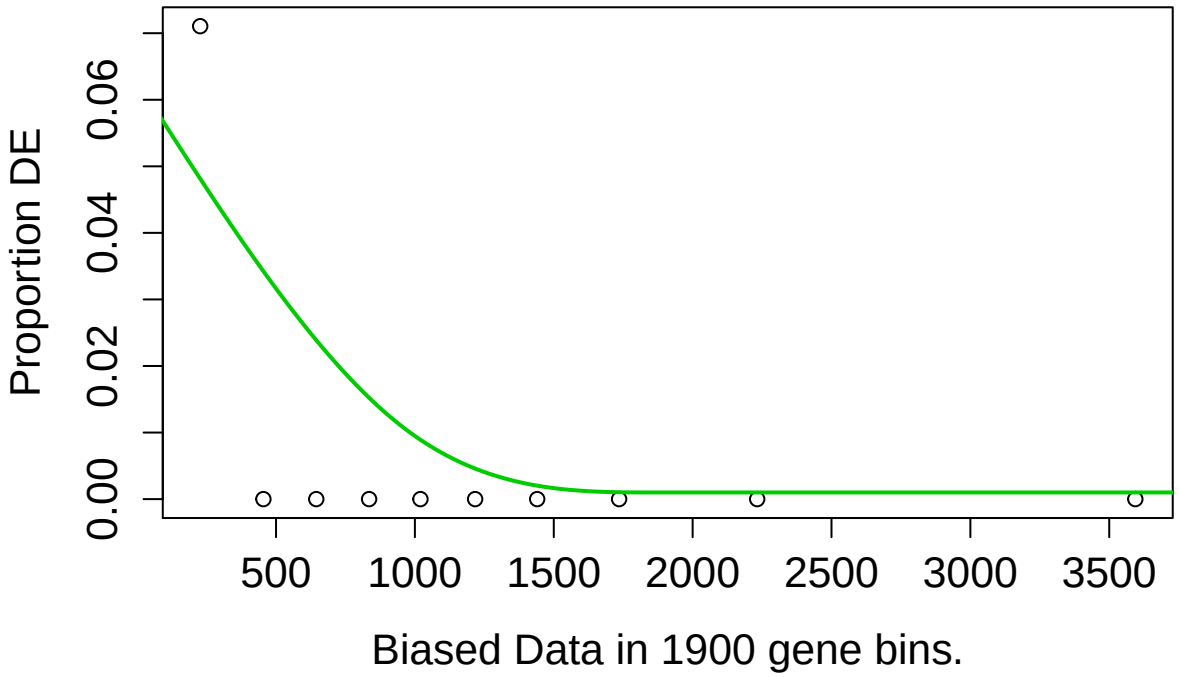



```
clusterGO(13)
```




```
## Using manually entered categories.
## For 3071 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```



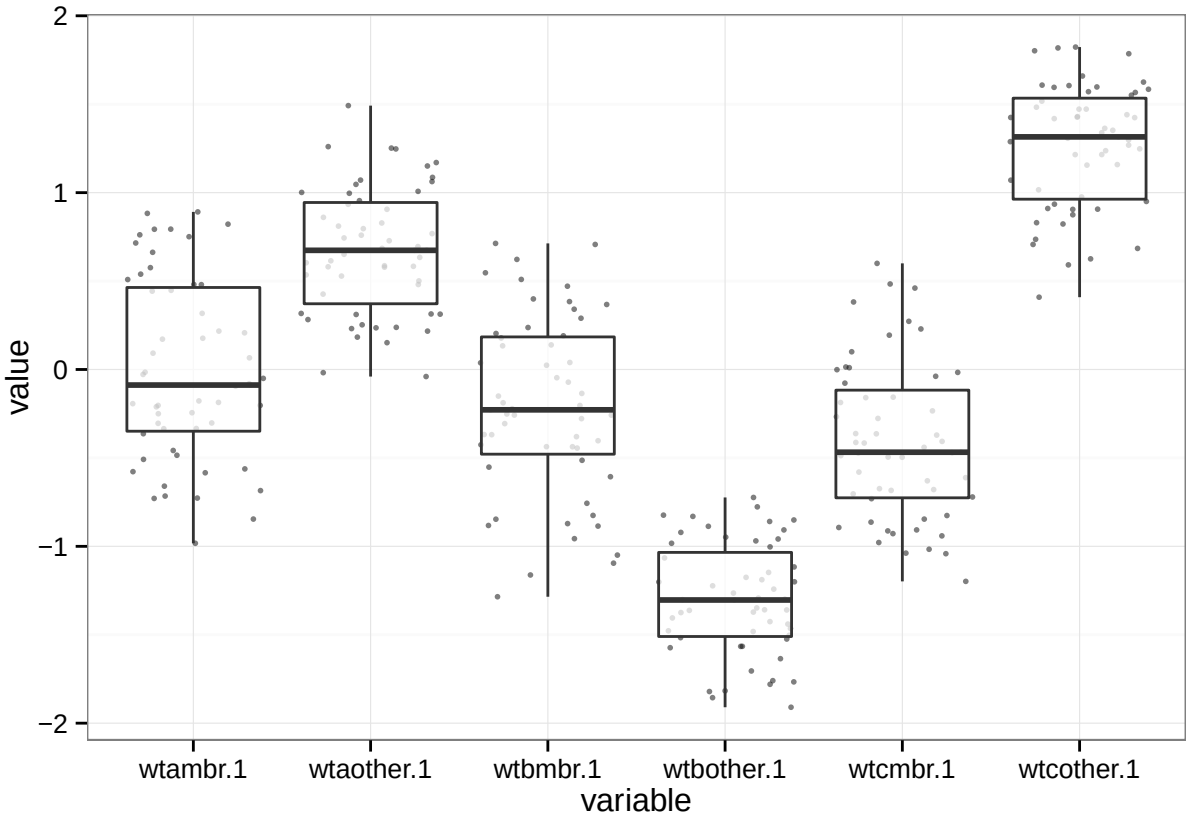

```
##      [,1]
```

```
y <- genesInClust(13, data.val2, annotation)
```

```
## [1] 125
```

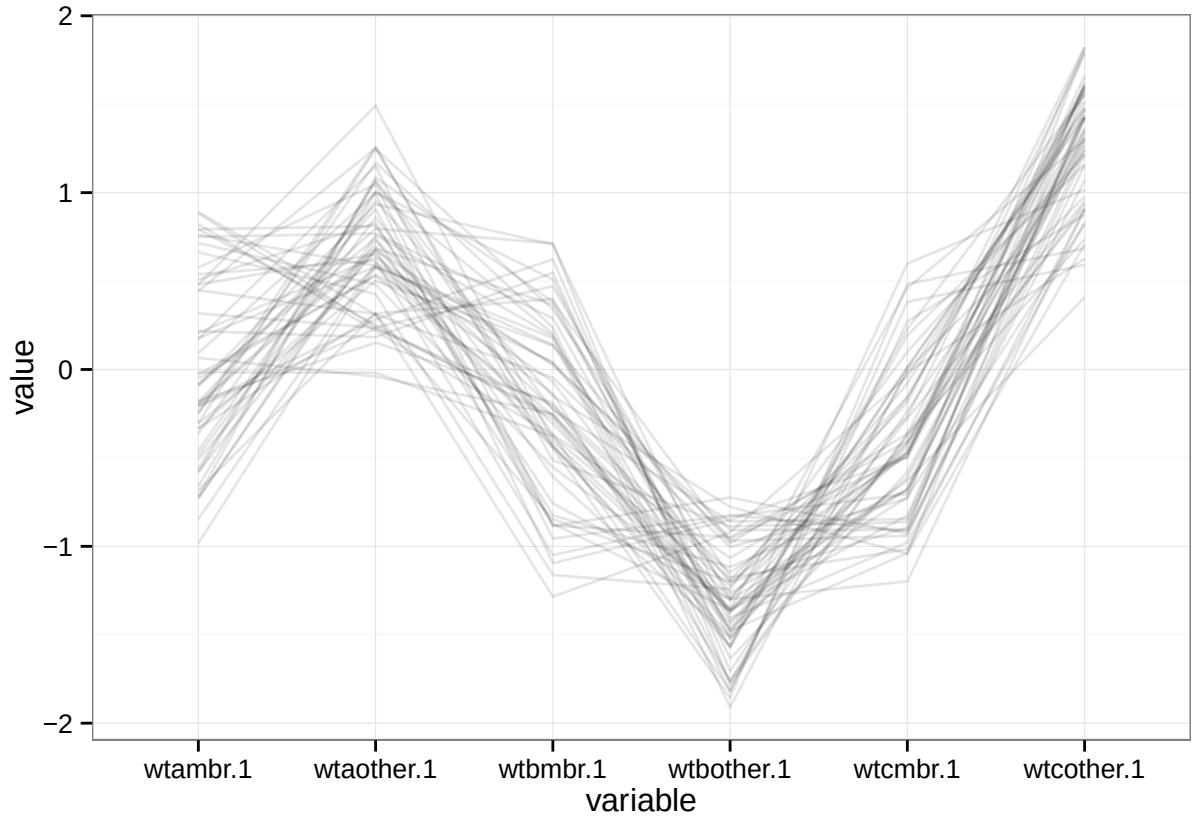
```
clusterVis(14)
```

```
## Using gene as id variables
```

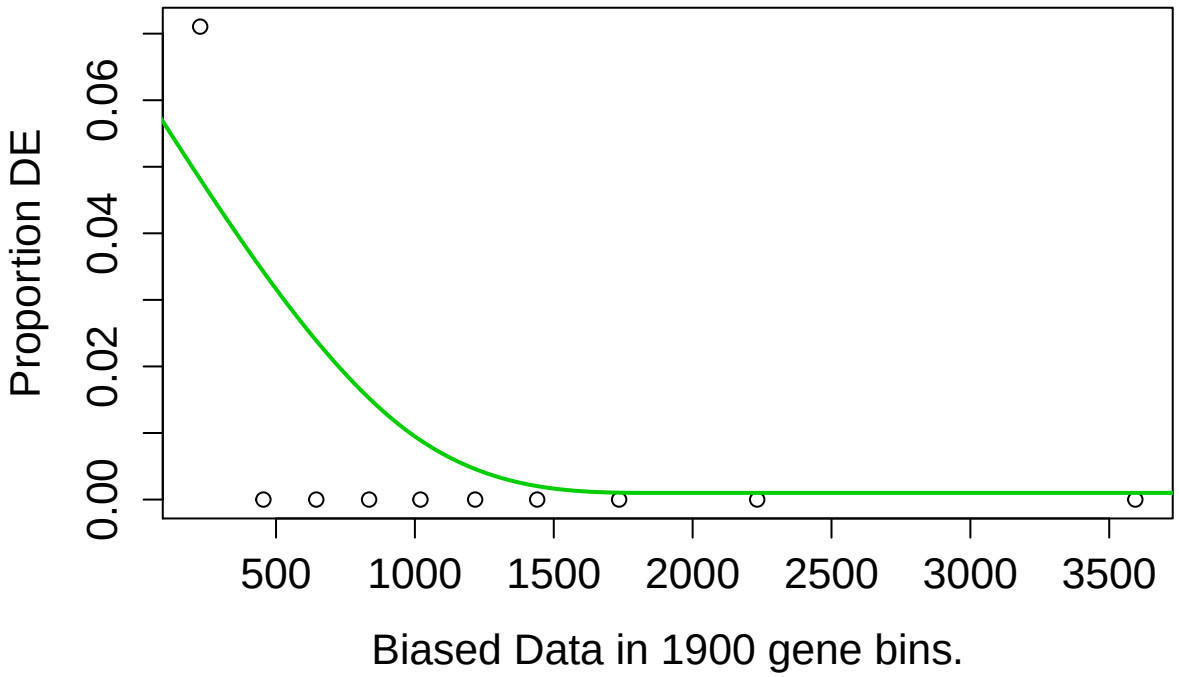



```
clusterVis_line(14)
```

```
## Using gene, curated as id variables
```

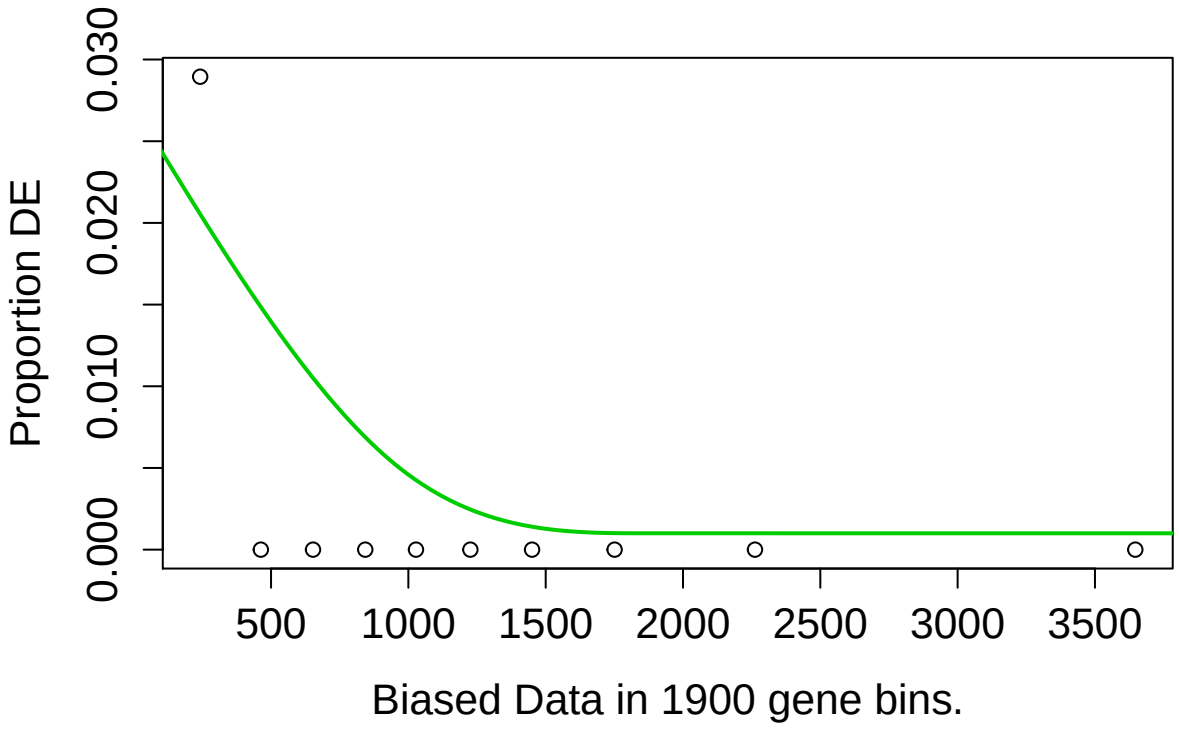



```
clusterGO(14)
```




```
## Using manually entered categories.
## For 2991 genes, we could not find any categories. These genes will be excluded.
## To force their use, please run with use_genes_without_cat=TRUE (see documentation).
```

```
## This was the default behavior for version 1.15.1 and earlier.
## Calculating the p-values...
```




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

```
y <- genesInClust(14, data.val2, annotation)
```

```
## [1] 125
```

Conclusions

I need to write something that looks at the statistical difference in tissue between each genotype. In order to do that I have to subset based on tissue and do a t-test? Do I need to correct for multiple testing? Ideally what information do I want from these clusters?

1. I want clusters that are enriched in leaf genes. Maybe I could just color special for leaf genes?
2. Significant differences between genotype at each tissue.

### How much intersect is there between my genes and the curated gene list?

```
length(intersect(data.val2$gene, genesOfInterest$gene) ) #77
```

```
## [1] 77
```

Only 51? Is that right? Double check. Are there only 51 leaf curated genes differentially expressed? Maybe I need to be looking at a larger subset. 25% co-efficient of variation. In `lcmSOM_analysis4_072914.Rmd`, I looked into this a bit more.

```
#From lcmSOM_analysis4_072914.Rmd
intersect(genesOfInterest$gene, countData$X) #There are only 217 leaf curated genes in the normalized r
intersect(genesOfInterest$gene, data.val.allGenes25$gene) # There are only 77 which are in the top 25%
#There are 51 which overlap with the DE genes.
```

I have to make a decision between