

`rnaseq_figure_plotter` (by Atsumi Ando, email ; a.ando@utexas.edu)

Usage; Rscript rnaseq_figure_plotter.r -i input_file -t bar -o output_file -g gene_list_file ... -c 5 -s 6

parameter of `rnaseq_figure_plotter`

HELP -h, --help show this help message and exit

required function

INPUT -i, --input input file name

TYPE -t, --type choose plot types (bar, box, density, density_fill, dot_color, dot_shape, heatmap, histogram, line, scatter, or violin)

general optional function

OUTPUT -o, --output default output; output file name

GENE -g, --gene file name of specific gene ID list; generate "output" _gene_selection.txt file

LOG2 -l, --log default 0; calculate log value (log2; 2, log10; 10, loge; e)

LOG2_NUMBER -lgn, --log_number default 0.000000001; add number to avoid -inf for log value

ZSCORE -zs, --zscore default off; apply Z-score transformation in gene (on or off). --log function should be 0 to apply --zscore function.

XAXIS -x, --xaxis default samples; choose x-axis (gene or sample)

ZAXIS -z, --zaxis default gene; choose fill, color, or shape (gene or sample)

COLOR -c, --color default 1; choose color type (1-10)

CUSTOM_COLOR -cst, -custom_color default None; customize color scales. Split colors by space. Example; red white blue green yellow

LETTER_SIZE -ls, --letter_size default 8 10; type text and title size of legend and axis, respectively. Split two number by space. Example; 20 24

FIGURE_SAVE_FORMAT -f, --figure_save_format default pdf; choose format of figures (eps, ps, tex (pictex), pdf, jpeg, tiff, png, bmp, svg)

optional parameter for individual plot types

STYLE -s, --style default 4; choose background of figures (1-7). This function works for any plots except heatmap.

LIMIT -lim, --limit default None; apply individual scale of "data". This function works for any plots except heatmap. Split two numbers(e.g. limit 0 to 200 -> type 0 200) by space. Negative number required double quotation marks such as "negative number". Example; 0 100/-1" 3

AXIS_CHANGE -a, --axis_change default off; flip axis in figures (on or off). This function works for any plots except heatmap.

LEGEND_POSITION -lp, --legend_position default right; choose legend position of figures (none, left, right, bottom, top, or two-element numeric vector). This function works for any plots except heatmap and scatter.

GEOM_POSITION -gp, --geom_position default 1; choose plot visualize types (geom position) from 1-4 in bar, density, and histogram

CLUSTER_SELECT -cs, --cluster_select default on on; apply column and row cluster function for heatmap (on or off). Column is first and row is second, split two factor(on or off) by space. Example; on off

SCATTER_SELECT -ss, --scatter_select default None; type column of two samples for comparison in dot plot. Split samples by space. Example; sample1 sample2

PLOT_SIZE -p, -plot_size default 7 7; type width and height of figure. Split two number by space. This function works for any plots except heatmap.

Example; 10 12

Input Data

-i, --input input_file

Input data should be samples for column and gene ID for row.

	sample1	sample2	sample3	sample4	sample5
geneA	1	3	5.5	7	2
geneB	100	267	55	79	62
geneC	0.3	0.65	9.5	0.87	2.1
geneD	205	356	78	67	2900
geneE	1001		3001		5500
geneF	2	2	2	2	2
geneG	0.01	0.03	0.5	0.07	0.02

Optional; if you need to extract specific genes, provide -g, --gene gene_id_list_file

-g, --gene
gene_id_list_file

Gene ID should be in first row and split by \n.

geneA
geneF
geneG

-g, --gene function provide "output" _gene_selection.txt file as followings.

"sample1"	"sample2"	"sample3"	"sample4"	"sample5"	
"geneA"	1	3	5.5	7	2
"geneF"	2	2	2	2	2
"geneG"	0.01	0.03	0.5	0.07	0.02

Dataframe generation

Optional; add log transformation (log2, log10, loge) or z-score calculation. To use (-zs, --zscore) function, (-l, --log) function requires to be off (0).

<u>Parameter</u>	<u>Setting</u>	<u>Description</u>
-l, --log	2, 10, or e	return log2, log10, or loge
-zs, --zscore	on	return z-score transformation

Once software finishes it provide dataframe of three columns, gene, data, and sample. Sample, data, and gene refer to sample name, gene expression value, and gene ID, respectively. Example of dataframe is following;

sample	data	gene
sample1	1.0000	geneA
sample1	100.0000	geneB
sample1	0.3000	geneC
sample1	205.0000	geneD
sample1	1001.0000	geneE
sample1	2.0000	geneF
sample1	0.0100	geneG

Once is successfully generate data frame it will show
"Complete dataframe generation! Dataframe generation time (seconds) ; time"

Axis selection

Default of x-axis and z-axis is sample and gene. You can change both axis by followings.

-x gene

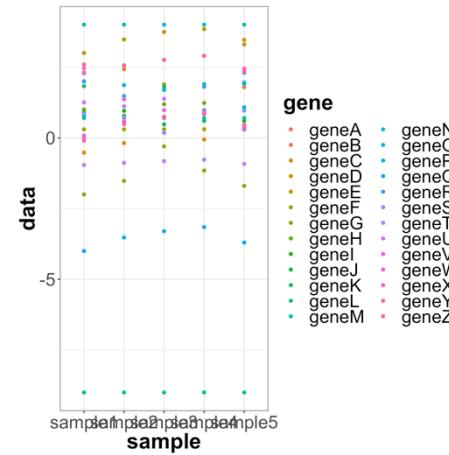
-z sample

You can modify x and z in the following table.

plots	x-axis	y-axis	color/shape
bar	x	data	z
box	x	data	x
density	data	density	x
dot_color	x	data	z
dot_shape	x	data	z
heatmap	sample	gene	
histogram	data	count	x
line	x	data	z
scatter			
violin	x	data	x

Rscript rnaseq_figure_plotter.r

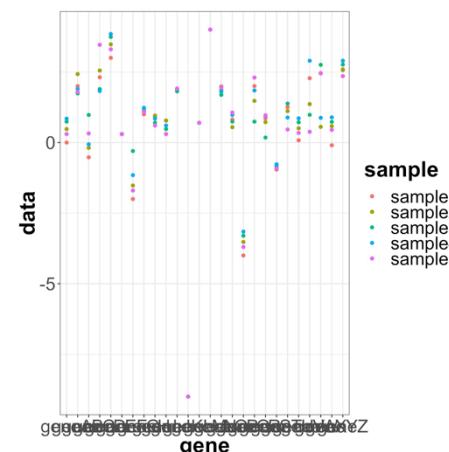
-i gene_expression_data.txt -t dot_color -l 10



Rscript rnaseq_figure_plotter.r

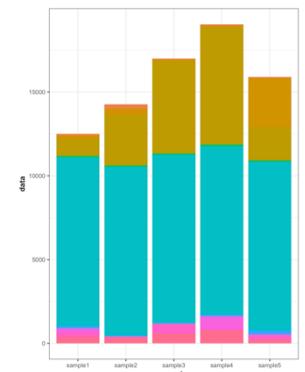
-i gene_expression_data.txt -t dot_color -l 10

-x gene -z sample

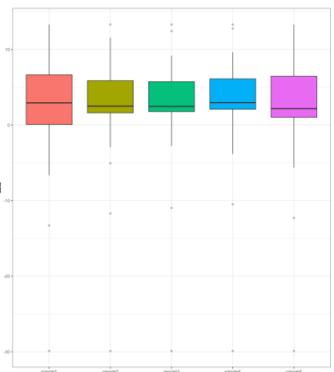


Plot types (Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt)

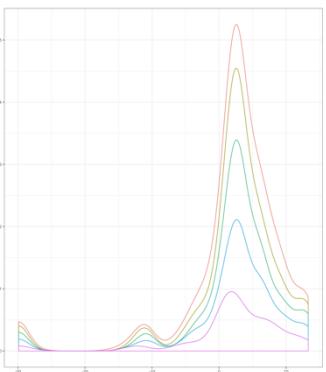
-t bar



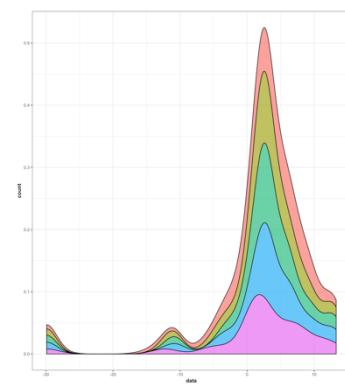
-t box



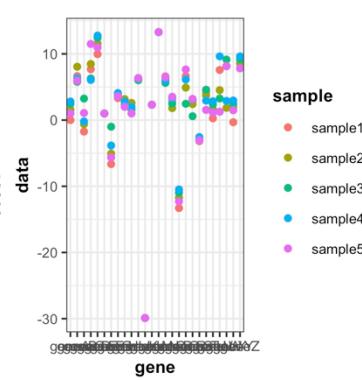
-t density



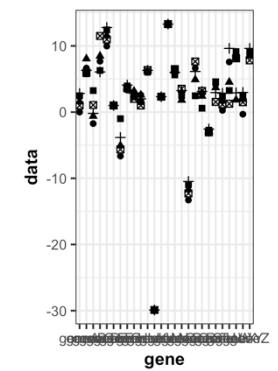
-t density_fill



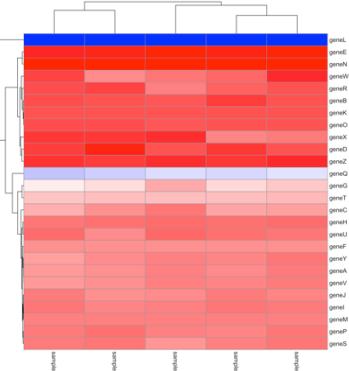
-t dot_color



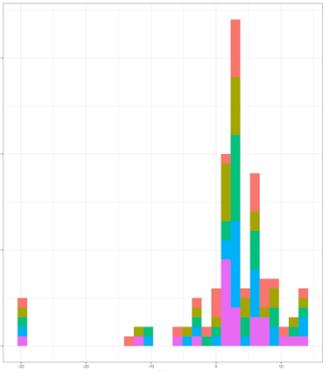
-t dot_shape



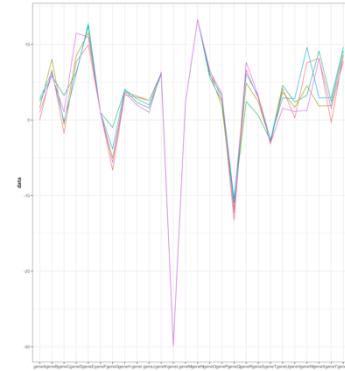
-t heatmap



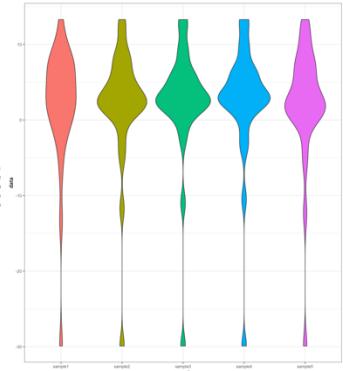
-t histogram



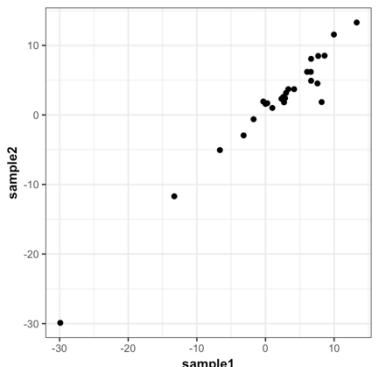
-t line



-t violin



-t scatter -ss sample1,sample2



Plot is generated in same directory and name as
“output”_“type”.“figure_save_format”

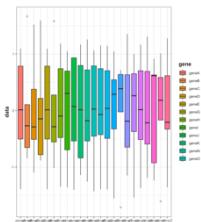
e.g. -o test -t bar -f jpeg

File in you directory; test_bar.jpeg

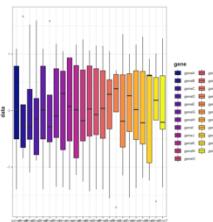
Color (Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t box/heatmap -x gene -zs on)

You can customize color by type **-cst**, **--custom_color (colors)**. Look example 4

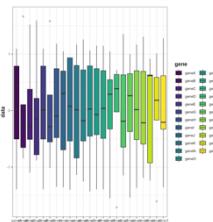
-c 1 (default)



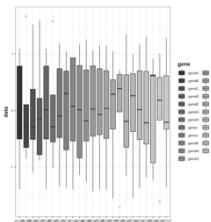
-c 2



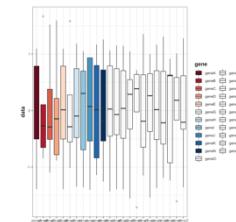
-c 3



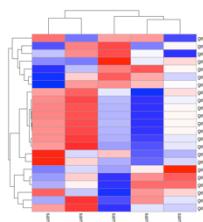
-c 4



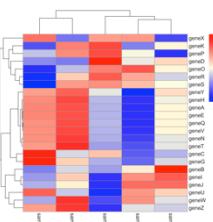
-c 5 (max9)



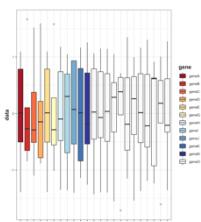
ggplot2



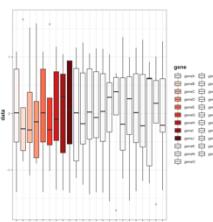
pheatmap



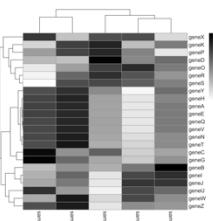
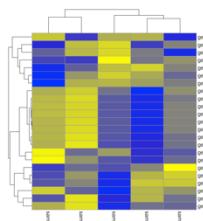
-c 6 (max 9)



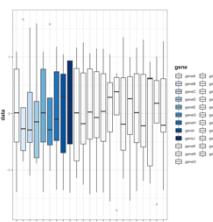
ggplot2



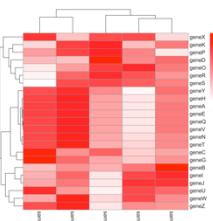
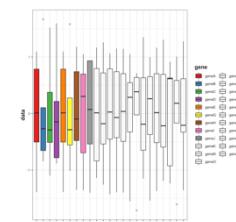
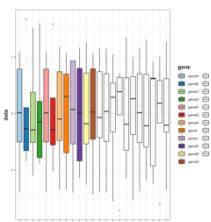
pheatmap



-c 8(max 9)



-c 9 (max 12)



-c 10 (max 9)

*max color number is for ggplot2 color

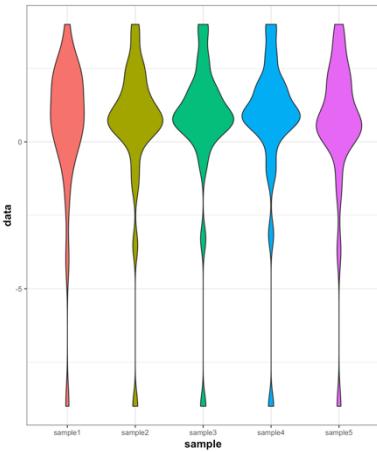
Letter and figure size/file format

Letter Default 8 10; type text and title size of legend and axis, respectively. Split two number by space. Example; 20 24.

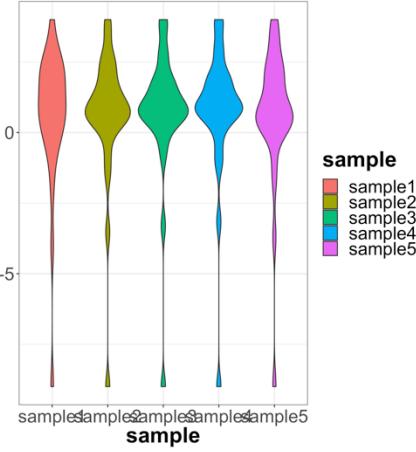
Figure Default 7 7; type width and height of figure. Split two number by space. This function works for any plots except heatmap. Example; 10 12

```
Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t violin -l 10
```

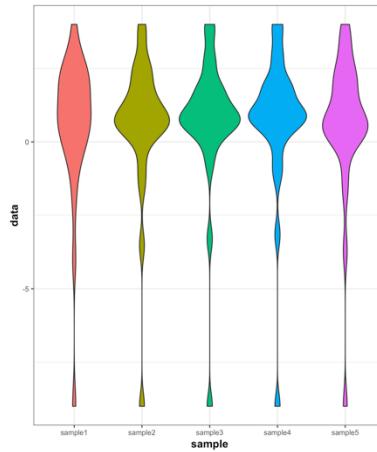
-ls 8 10 (default)



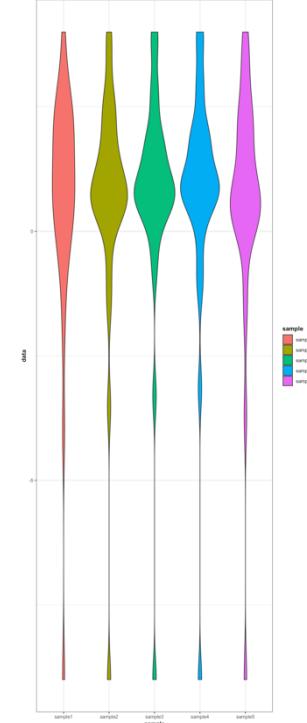
-ls 20 24



-p 7 7 (default)



-p 7 17

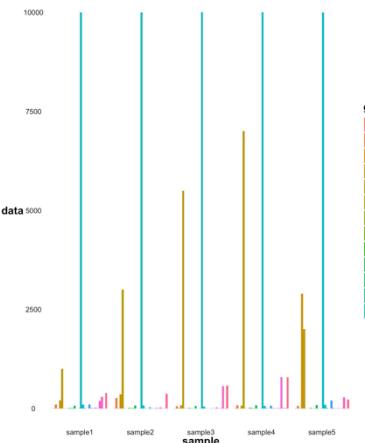


Default is pdf, you can also choose eps, ps, tex (pictex), pdf, jpeg, tiff, png, bmp, svg.
Example of usage; save figure in jpeg by (-f jpeg).

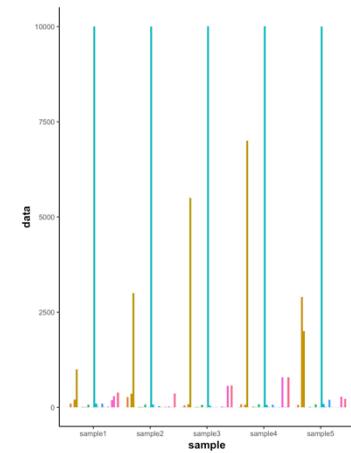
Style (Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t bar -gp 2 -f jpeg)

*This function is for every plots excepts heatmap.

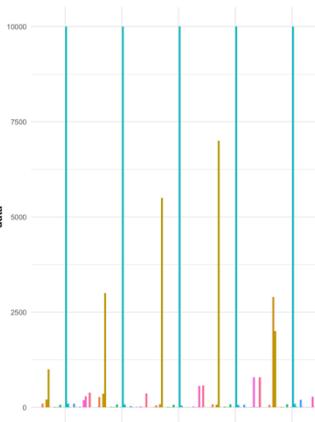
-s 1



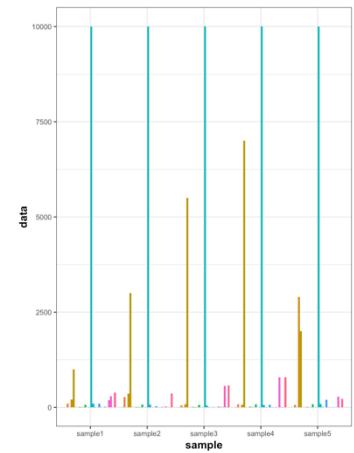
-s 2



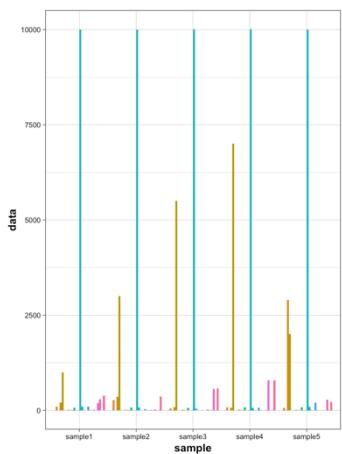
-s 3



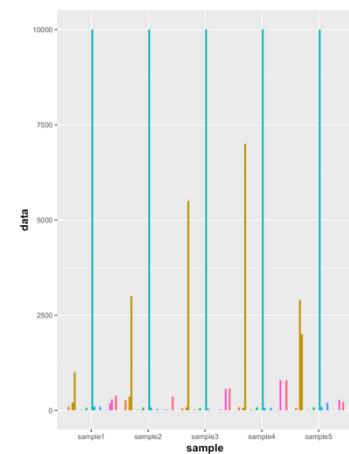
-s 4 (default)



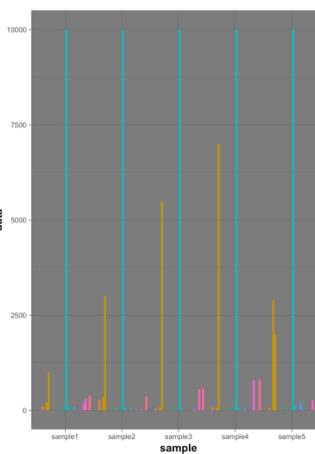
-s 5



-s 6



-s 7



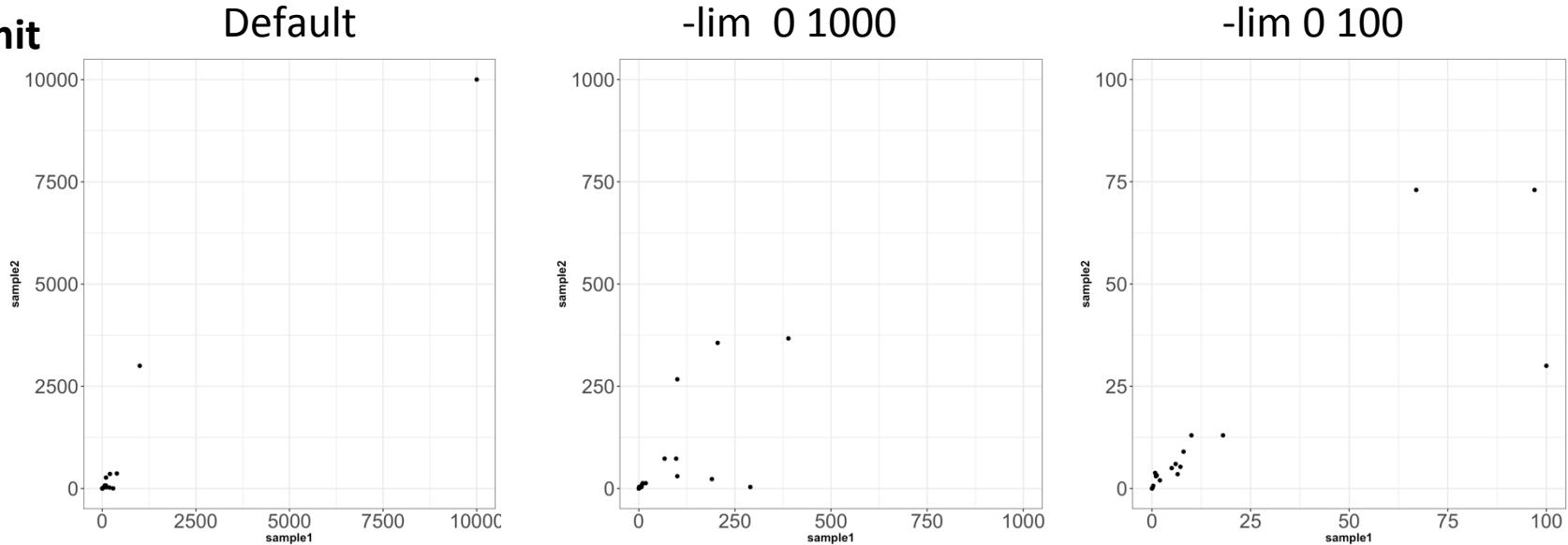
Data limit and axis change

Data limit (Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t scatter -ss sample1 sample2)

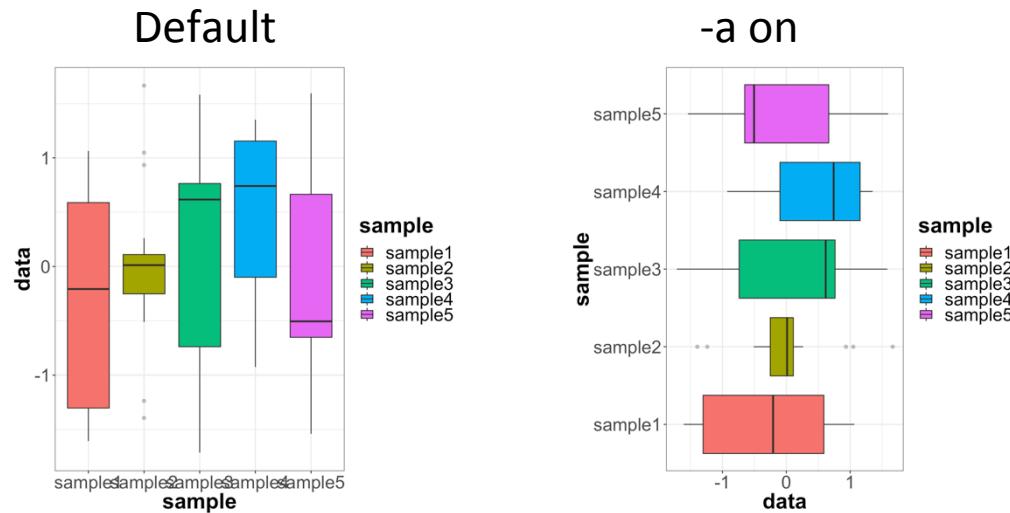
Axis change(Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t box -zs on)

*This function is for every plots excepts heatmap.

Limit



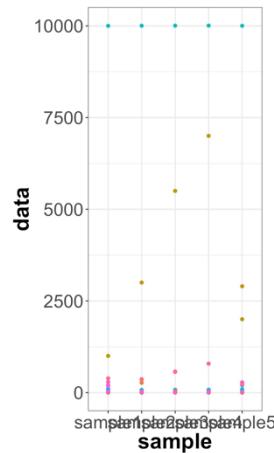
Axis_change



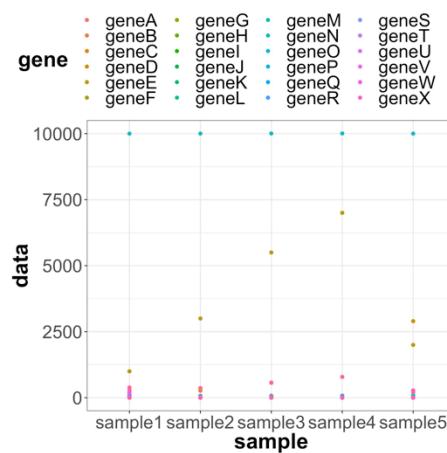
Legend position (Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t dot_color -ls 20 24)

*This function is for every plots excepts heatmap and scatter.

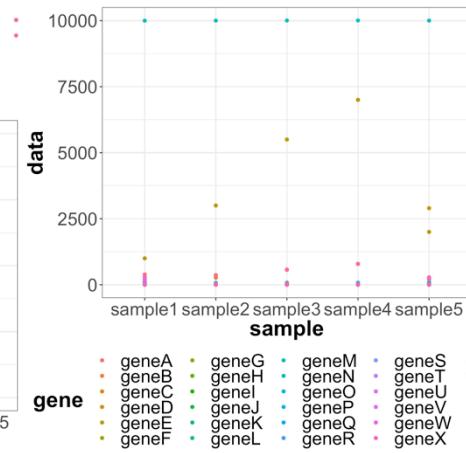
-lp right (default)



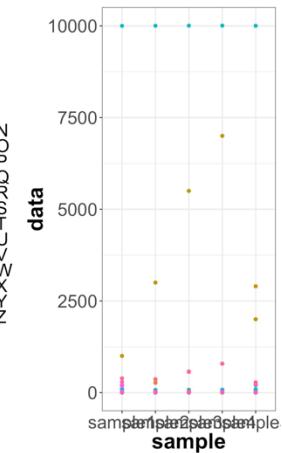
-lp top



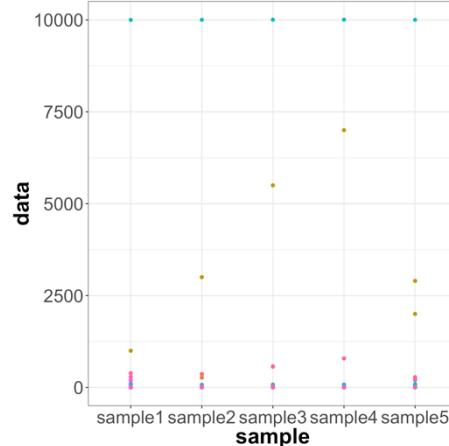
-lp bottom



-lp left



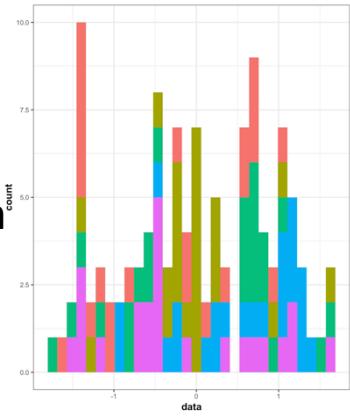
-lp none



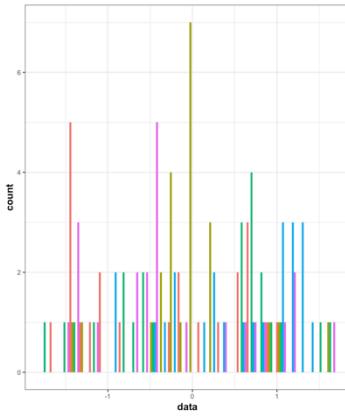
Geom_position (Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t histogram/density -zs on)

*This function is for bar, density, and histogram.

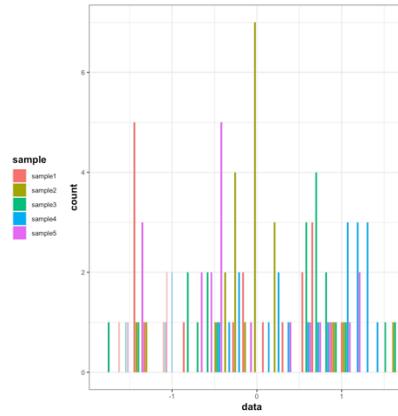
-pg 1 (default)



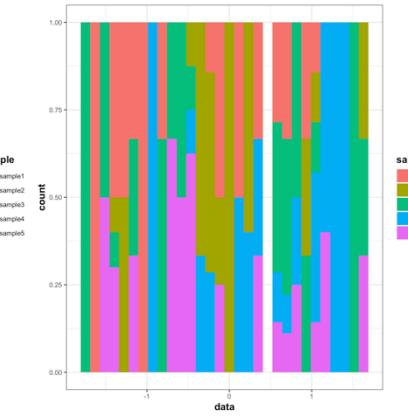
-pg 2



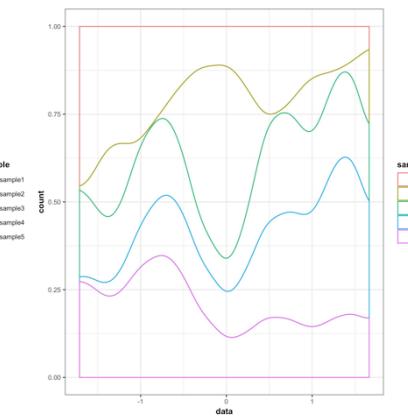
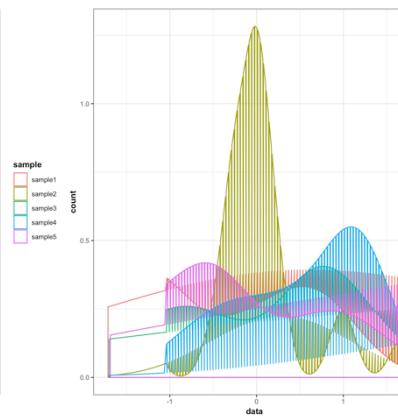
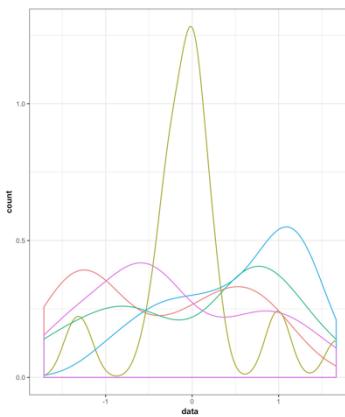
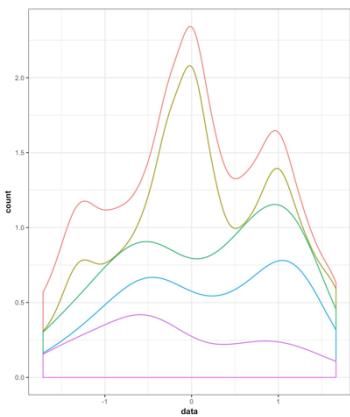
-pg 3



-pg 4



histogram

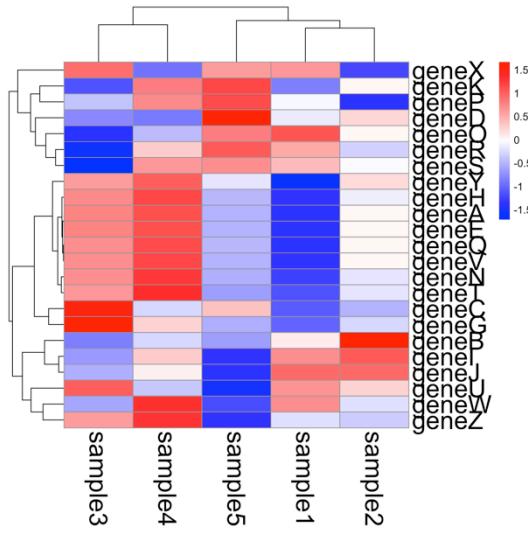


density

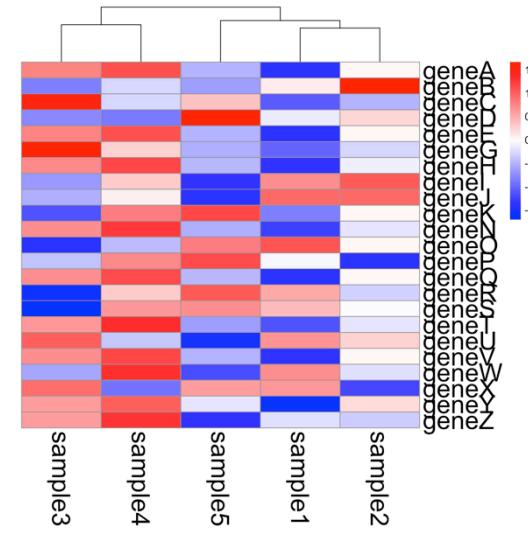
Cluster (Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t heatmap -zs on -ls 20 24)

*This function is only for heatmap.

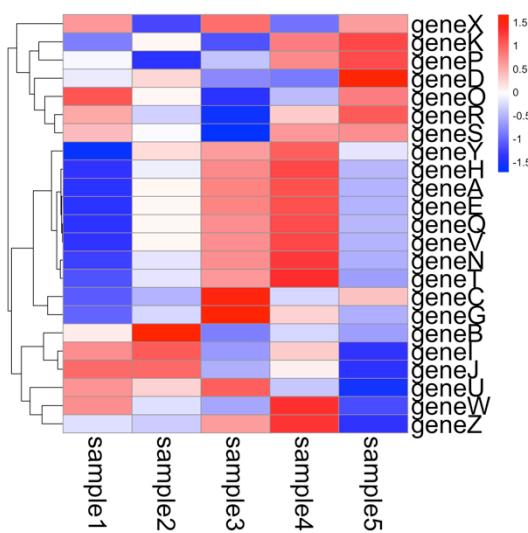
-cs on on (default)



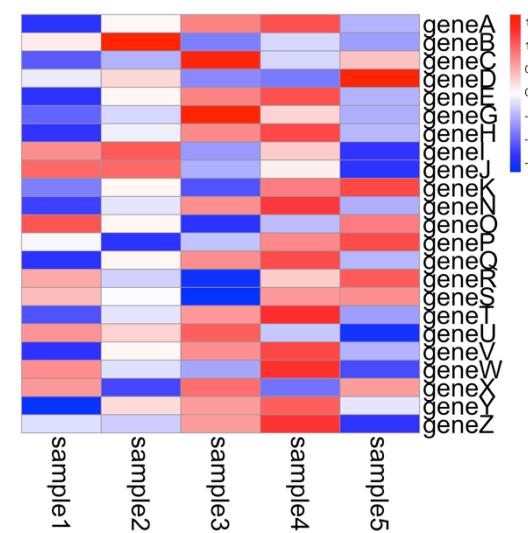
-cs on off



-cs off on



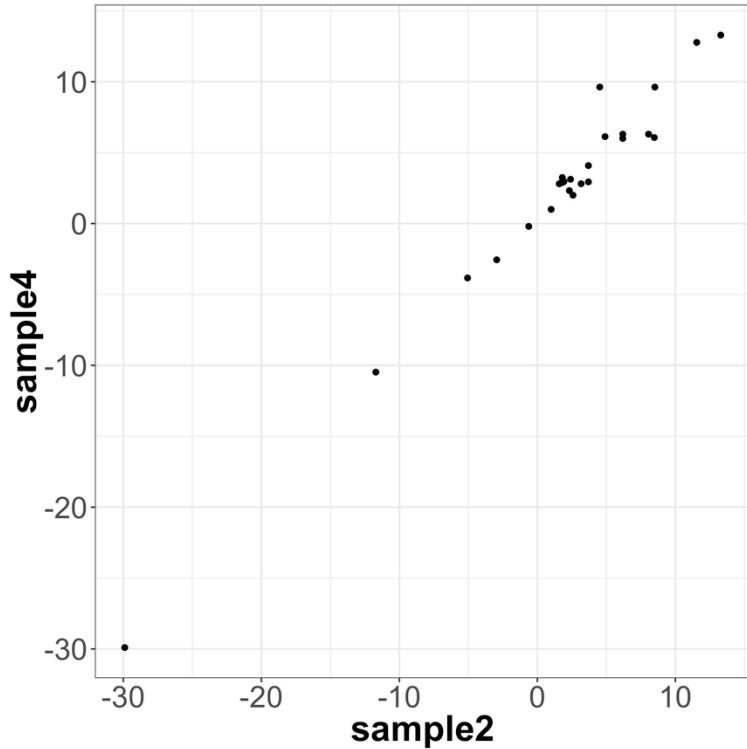
-cs off off



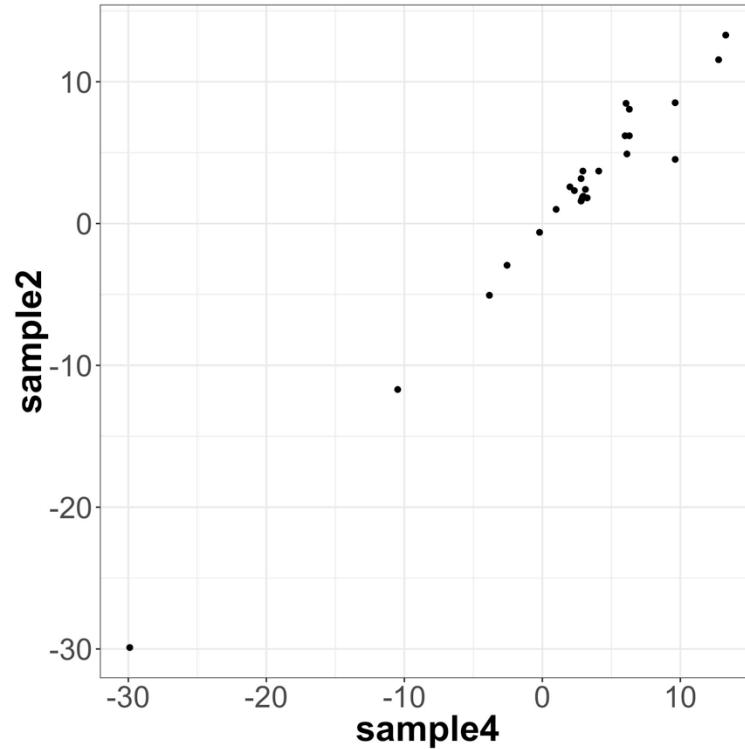
Scatter (Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t scatter -ls 20 24 -l e)

) *This function is only for scatter plot.

-ss sample2 sample4



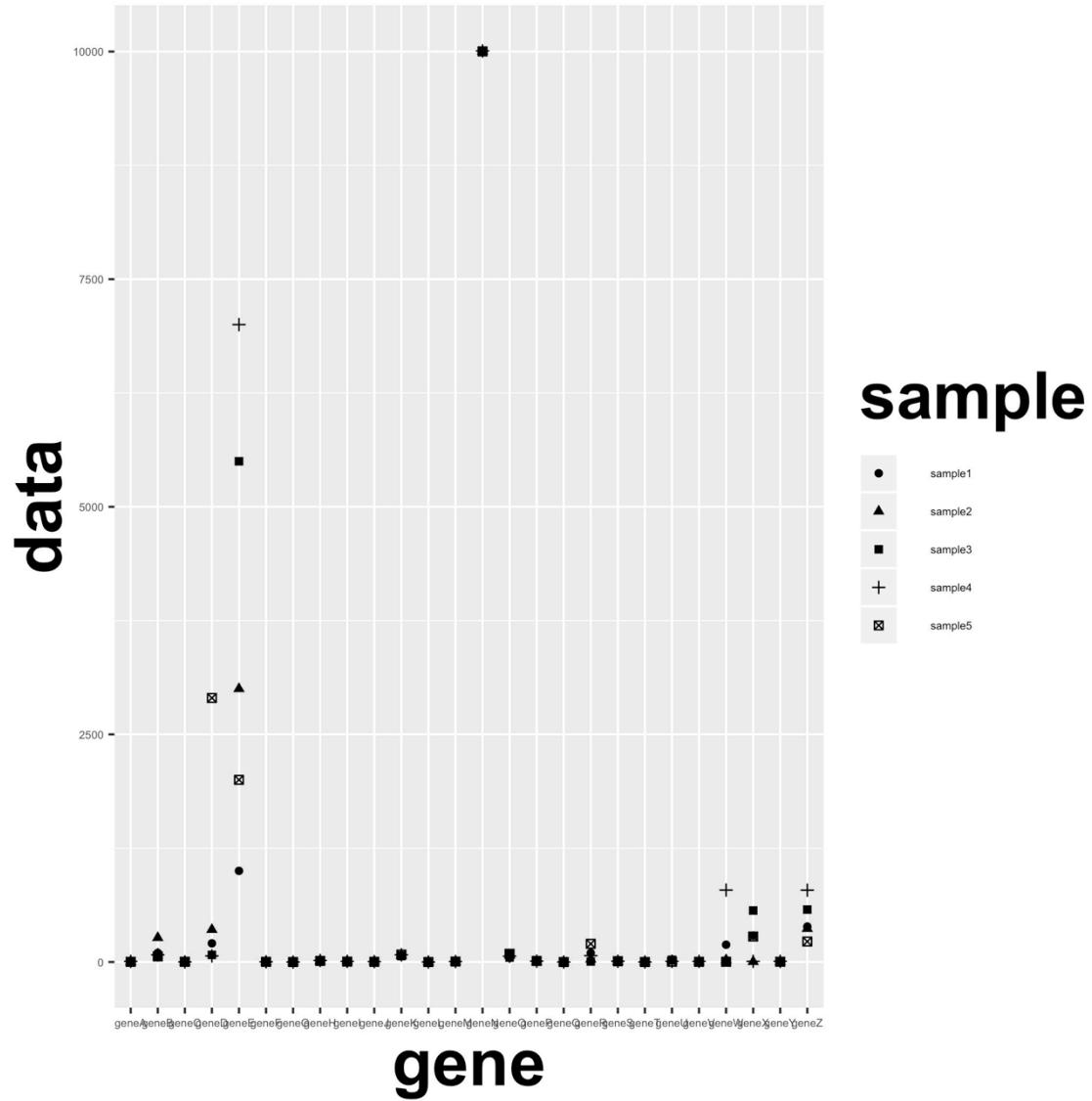
-ss sample4 sample2



Once it successfully generate figure it will show
“Plotting now! Total operation time (second); time” and provide figure as
“output”_“type”.“figure_save_format” file (e.g. output_bar.pdf)

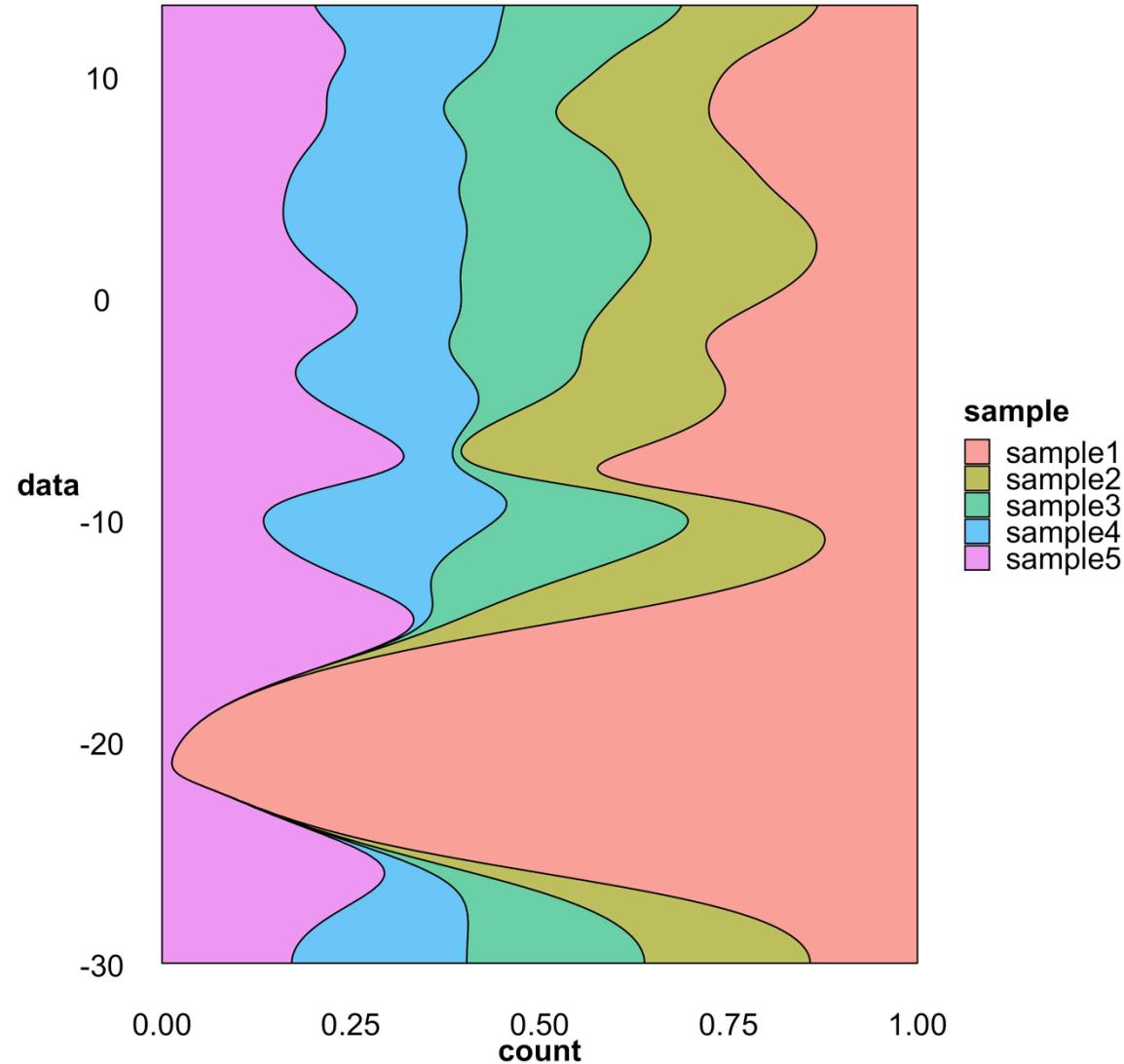
Example1

```
Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t dot_shape -o sample_data  
-x gene -z sample -s 6 -ls 5 30 -f tiff
```



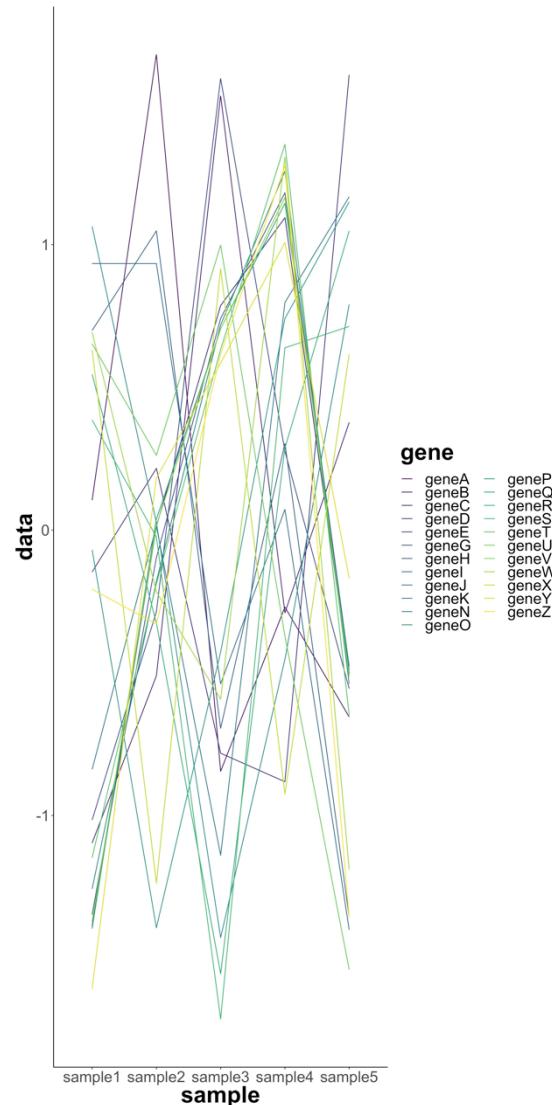
Example2

```
Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t density_fill -o sample_test  
-l e -s 1 -a on -gp 4 -ls 20 20
```



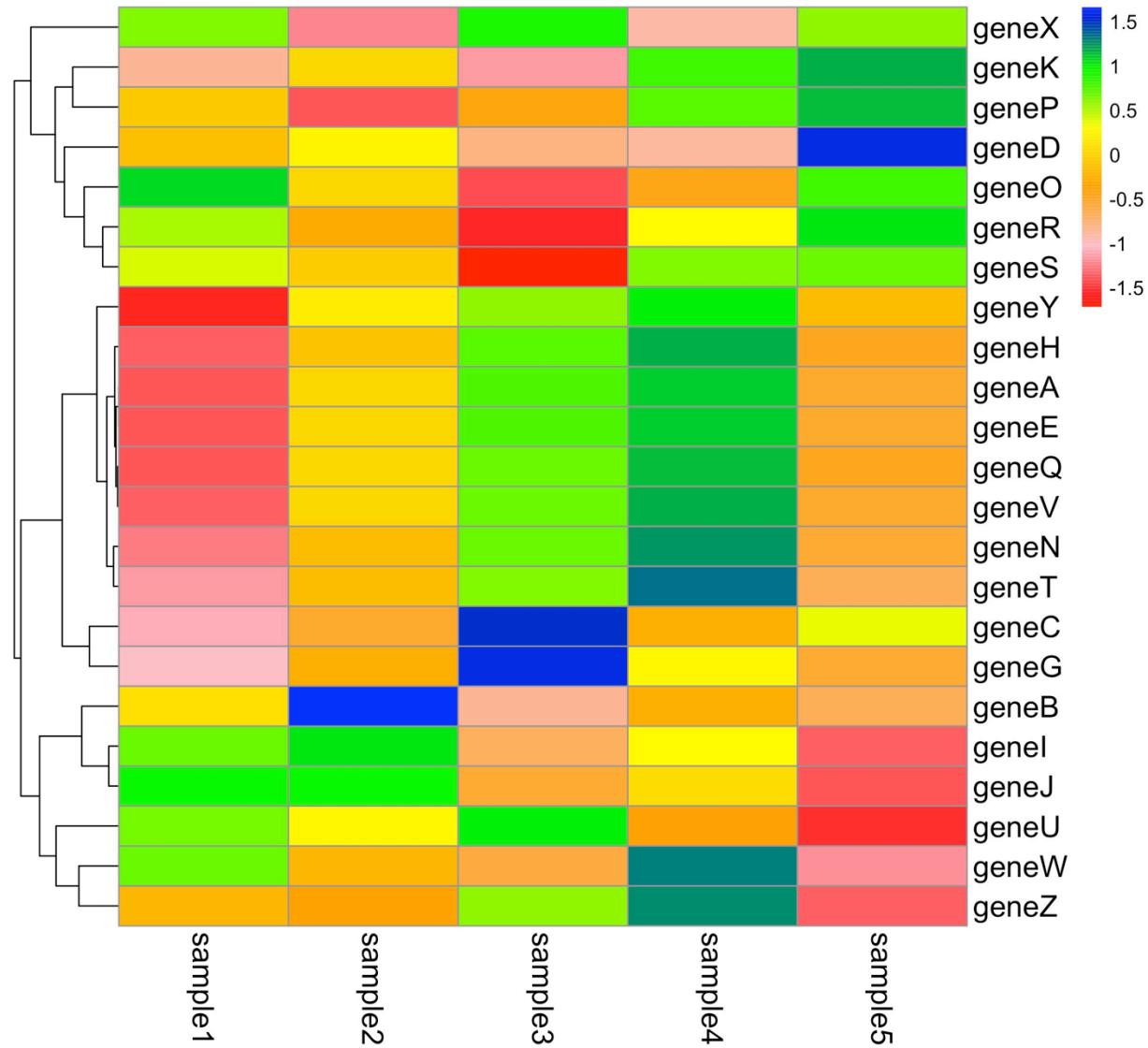
Example3

```
Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t line -zs on -s 2 -c 3 -ls 20 30  
-f png -p 10 20
```



Example4

```
Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t heatmap -o sample_heat  
-zs on -ls 15 15 -cs off on -cst red pink orange yellow green blue #custom color
```



Example5

```
Rscript rnaseq_figure_plotter.r -i gene_expression_data.txt -t bar -g gene_list.txt -l 10  
-c 6 -ls 15 15 -a on -gp 2 -lim "-1" 1.75
```

