

熱圖 Heatmap (矩陣視覺化)

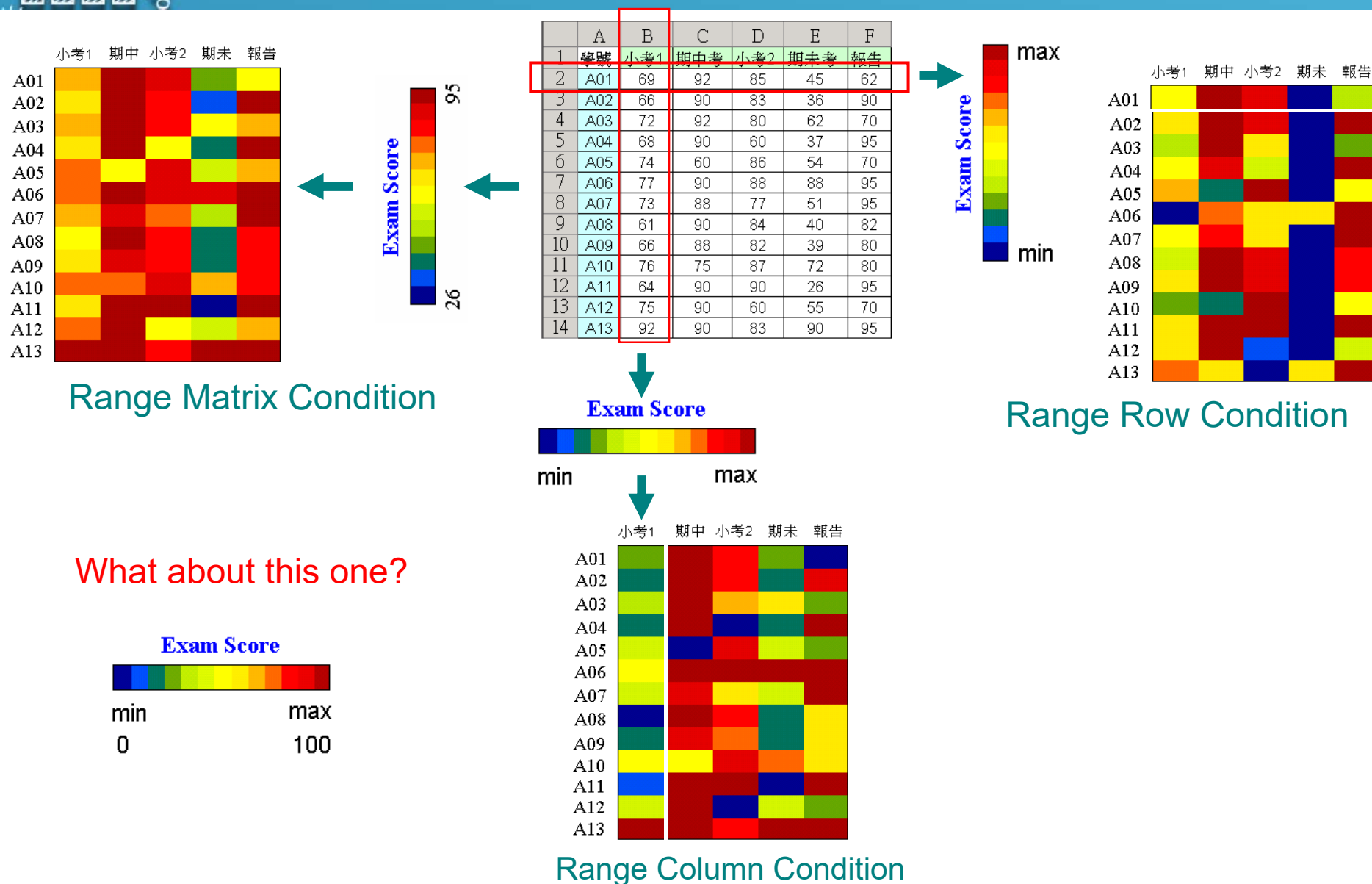
吳漢銘

國立政治大學 統計學系



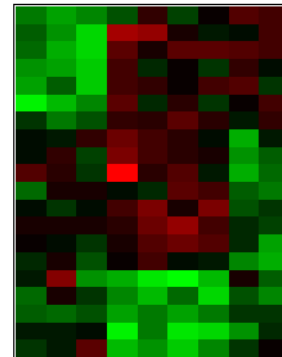
<https://hmwu.idv.tw>

Heat Map: Data Image, Matrix Visualization

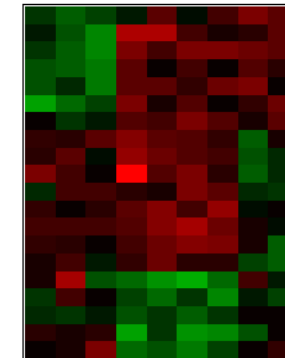


Heat Map: Display Conditions

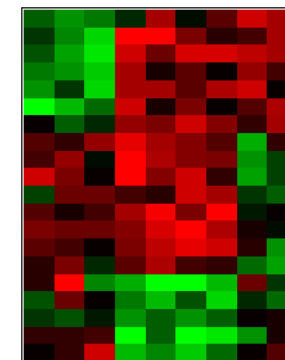
	A	B	C	D	E	F	G	H	I
1	-1.37	-2.30	-1.80	-0.55	2.45	-0.13	1.49	3.03	2.48
2	-0.68	-2.11	-3.42	4.67	4.57	1.75	0.61	0.92	2.52
3	-1.19	-2.49	-3.66	3.14	1.70	3.29	3.33	2.92	2.48
4	-1.93	-2.28	-3.16	2.51	0.32	1.49	0.21	2.20	1.03
5	-2.21	-0.79	-3.29	2.55	2.44	1.45	2.68	3.03	0.19
6	-4.14	-2.91	-1.64	3.21	0.37	1.93	0.14	1.27	2.67
7	0.21	-1.36	-0.44	2.22	1.85	3.11	2.03	0.67	2.40
8	1.13	0.79	2.25	3.65	2.52	2.09	1.13	-2.59	0.67
9	0.95	2.33	-0.07	3.89	2.72	2.13	1.75	-2.17	-0.90
10	3.04	1.85	0.21	7.07	2.01	3.05	0.78	-2.58	-1.04
11	-1.02	1.65	1.53	0.95	0.60	3.12	2.52	-0.77	-1.40
12	1.21	0.24	1.04	2.50	3.69	1.81	3.98	-0.33	0.11
13	1.74	1.60	1.70	2.02	3.45	4.46	2.69	0.41	-0.09
14	1.34	1.06	0.06	1.81	2.90	3.64	3.04	0.49	-2.33
15	0.57	1.81	-0.47	1.40	2.70	0.99	0.82	-1.61	-2.56
16	0.61	4.22	-2.03	-2.61	-4.00	-4.64	-2.92	1.55	-0.71
17	-1.13	1.64	0.01	-1.77	-2.85	-1.24	-3.41	-0.59	-1.64
18	-0.88	-1.17	-0.41	-2.20	-1.30	-2.37	-1.41	0.08	0.25
19	0.75	0.66	1.04	-4.26	-1.41	-3.99	-3.53	-2.17	0.34
20	0.15	0.68	3.18	-2.86	-2.01	-3.18	-1.58	0.10	1.28



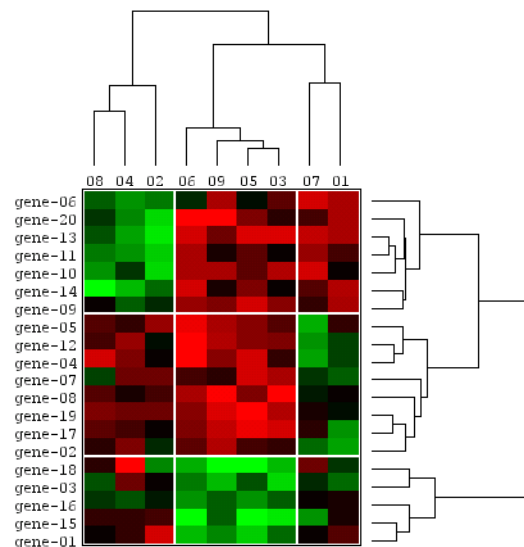
max: 7.07
7.07
5.22
3.37
1.53
-0.32
-2.17
-4.64
min: -4.64



max: 7.07
7.07
4.84
2.61
0.0
-1.86
-4.09
-7.07
min: -7.07



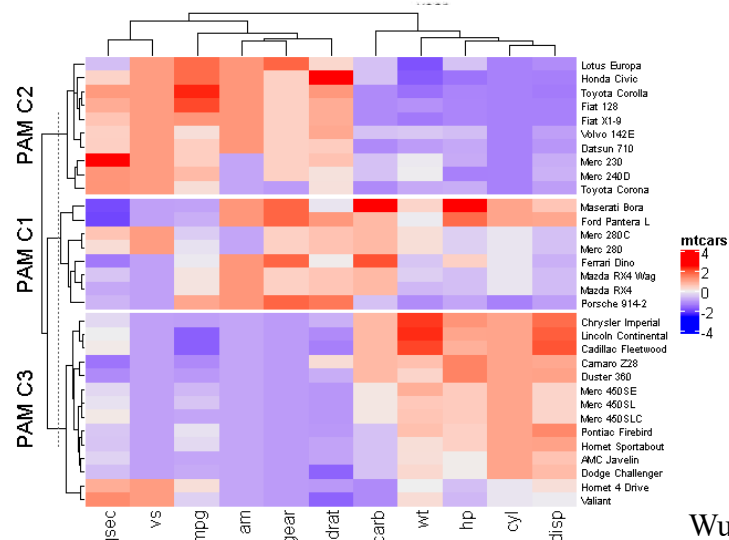
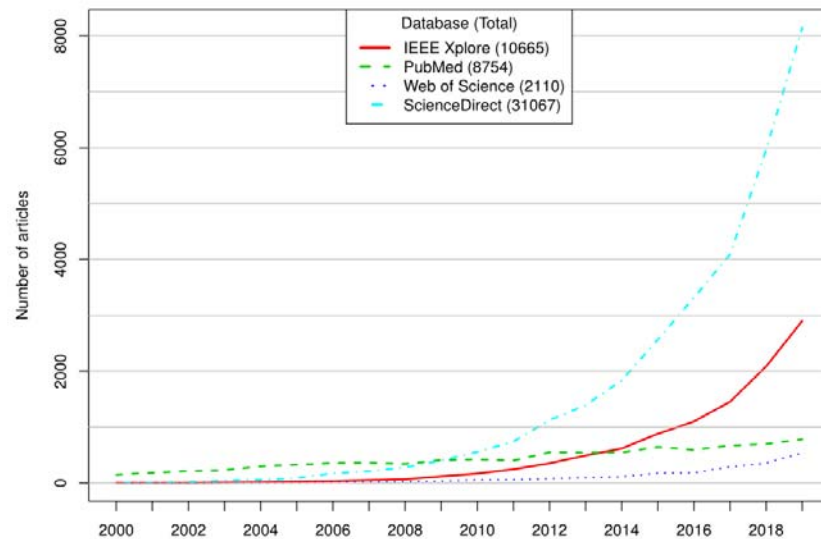
max: 7.07
7.07
4.0
2.81
1.51
0.0
-1.08
-2.37
-4.0
min: -7.07



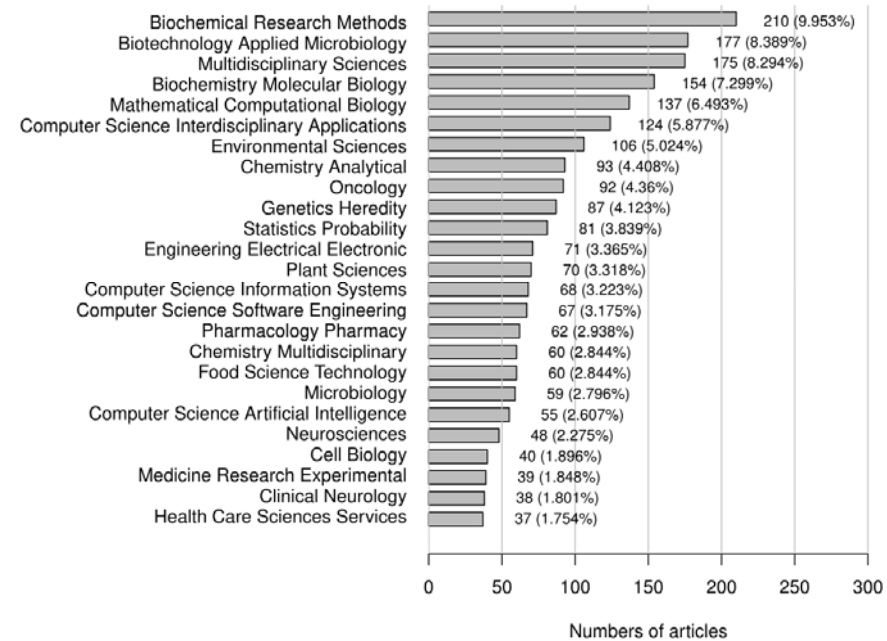
Center Matrix Condition

Heatmap

Heatmaps researches and applications from year 2000 to 2019



Top 25 heatmap researches and applications domains (WOS).



Wu, H. M., and Chen, C.H.* (2020), Matrix visualization: a review and perspective, WIRES Computational Statistics.

Table 1: R packages for static MV/heatmaps.

Packages	Description	Reference
Autoimage	Multiple heatmaps for projected coordinates	French (2017)
ComplexHeatmap	Make complex heatmaps	Gu <i>et al.</i> (2016)
corrplot	Visualization of a correlation matrix	Wei <i>et al.</i> (2017)
fheatmap	Fantastic heatmap	Tumulu and Sindiri (2015)
heatmap3	An improved heatmap package	Zhao <i>et al.</i> (2014)
heatmap.plus	Heatmap with more sensible behavior	Day (2015)
Heatplus	Heatmaps with row and/or column covariates and colored clusters	Ploner (2020)
NeatMap	Nonclustering heatmap alternatives in R	Rajaram and Oono (2010)
mcheatmaps	Multiple matrices heatmap visualization	Chenard and Najmanovich (2014)
pheatmap	Pretty heatmaps	Kolde (2015)
superheat	A graphical tool for exploring complex datasets using heatmaps	Barter and Yu (2018)

Table 2: R packages and R-based tools for interactive and/or web-based MV/heatmaps.

Packages	Description	Reference
clustvis	A web tool for visualizing clustering of multivariate data	Metsalu and Vilo (2015)
d3heatmap	Interactive heatmaps using <code>htmlwidgets</code> and <code>D3.js</code>	Cheng and Galili (2018)
gapmap	Drawing gapped cluster heatmaps with <code>ggplot2</code>	Sakai (2015)
heatmaply	Interactive cluster heatmaps using <code>plotly</code>	Galili <i>et al.</i> (2018)
Heatmapper	Web-enabled heat mapping for all	Babicki <i>et al.</i> (2016)
iheatmapr	Interactive, complex heatmaps	Schep and Kummerfeld (2017)
shinyheatmap	Ultrafast low-memory heatmap web interface for big data genomics	Khomtchouk <i>et al.</i> (2017)

Wu, H. M., and Chen, C.H.* (2020), Matrix visualization: a review and perspective, WIREs Computational Statistics.

Display of Genome-Wide Expression Patterns

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 14863–14868, December 1998
Genetics

Cluster analysis and display of genome-wide expression patterns

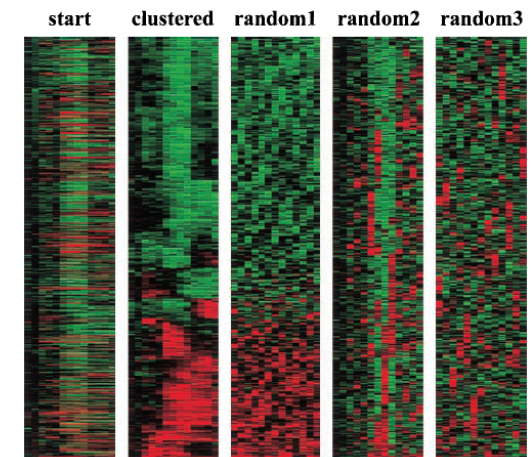
MICHAEL B. EISEN*, PAUL T. SPELLMAN*, PATRICK O. BROWN†, AND DAVID BOTSTEIN*‡

Software: Cluster and TreeView

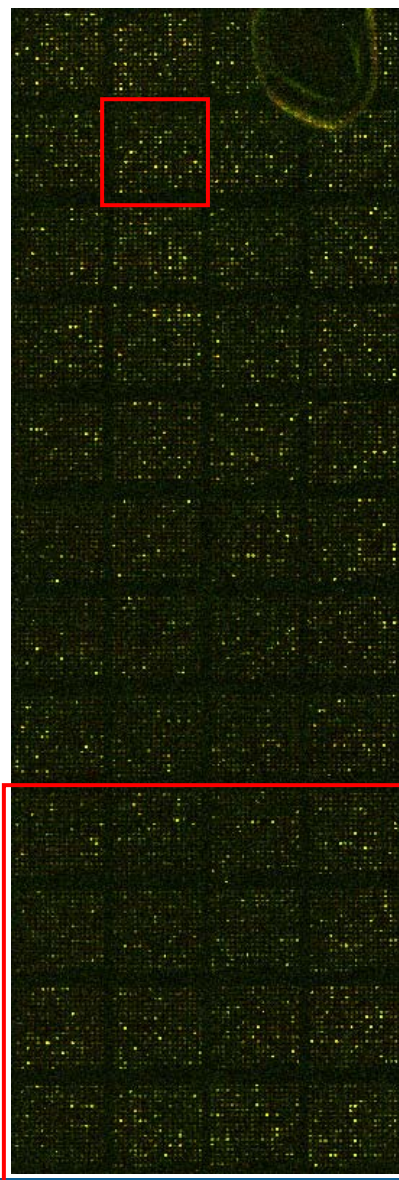
FIG. 1. Clustered display of data from time course of serum stimulation of primary human fibroblasts. Experimental details are described elsewhere (11). Briefly, foreskin fibroblasts were grown in culture and were deprived of serum for 48 hr. Serum was added back and samples taken at time 0, 15 min, 30 min, 1 hr, 2 hr, 3 hr, 4 hr, 8 hr, 12 hr, 16 hr, 20 hr, 24 hr. The final datapoint was from a separate unsynchronized sample. Data were measured by using a cDNA microarray with elements representing approximately 8,600 distinct

human genes. All measurements are relative to time 0. Genes were selected for this analysis if their expression level deviated from time 0 by at least a factor of 3.0 in at least 2 time points. The dendrogram and colored image were produced as described in the text; the color scale ranges from saturated green for log ratios -3.0 and below to saturated red for log ratios 3.0 and above. Each gene is represented by a single row of colored boxes; each time point is represented by a single column. Five separate clusters are indicated by colored bars and by identical coloring of the corresponding region of the dendrogram. As described in detail in ref. 11, the sequence-verified named genes in these clusters contain multiple genes involved in (A) cholesterol biosynthesis, (B) the cell cycle, (C) the immediate-early response, (D) signaling and angiogenesis, and (E) wound healing and tissue remodeling. These clusters also contain named genes not involved in these processes and numerous uncharacterized genes. A larger version of this image, with gene names, is available at <http://rana.stanford.edu/clustering/serum.html>.

FIG. 3. To demonstrate the biological origins of patterns seen in Figs. 1 and 2, data from Fig. 1 were clustered by using methods described here before and after random permutation within rows (random 1), within columns (random 2), and both (random 3).



Applications: Array Image



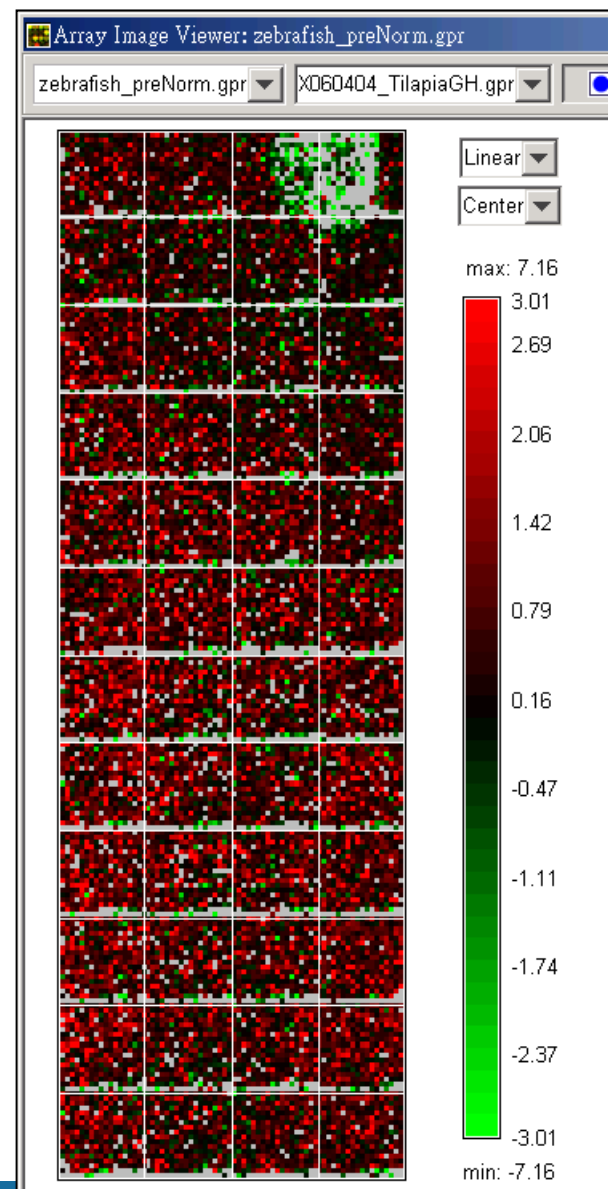
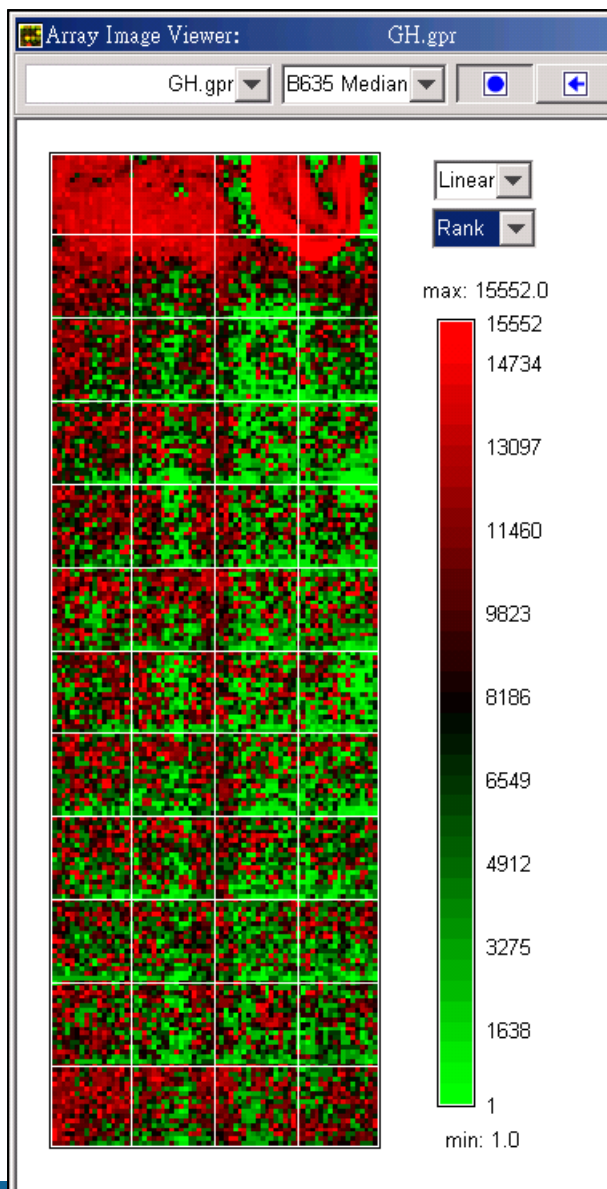
Blocks:
12 by 4

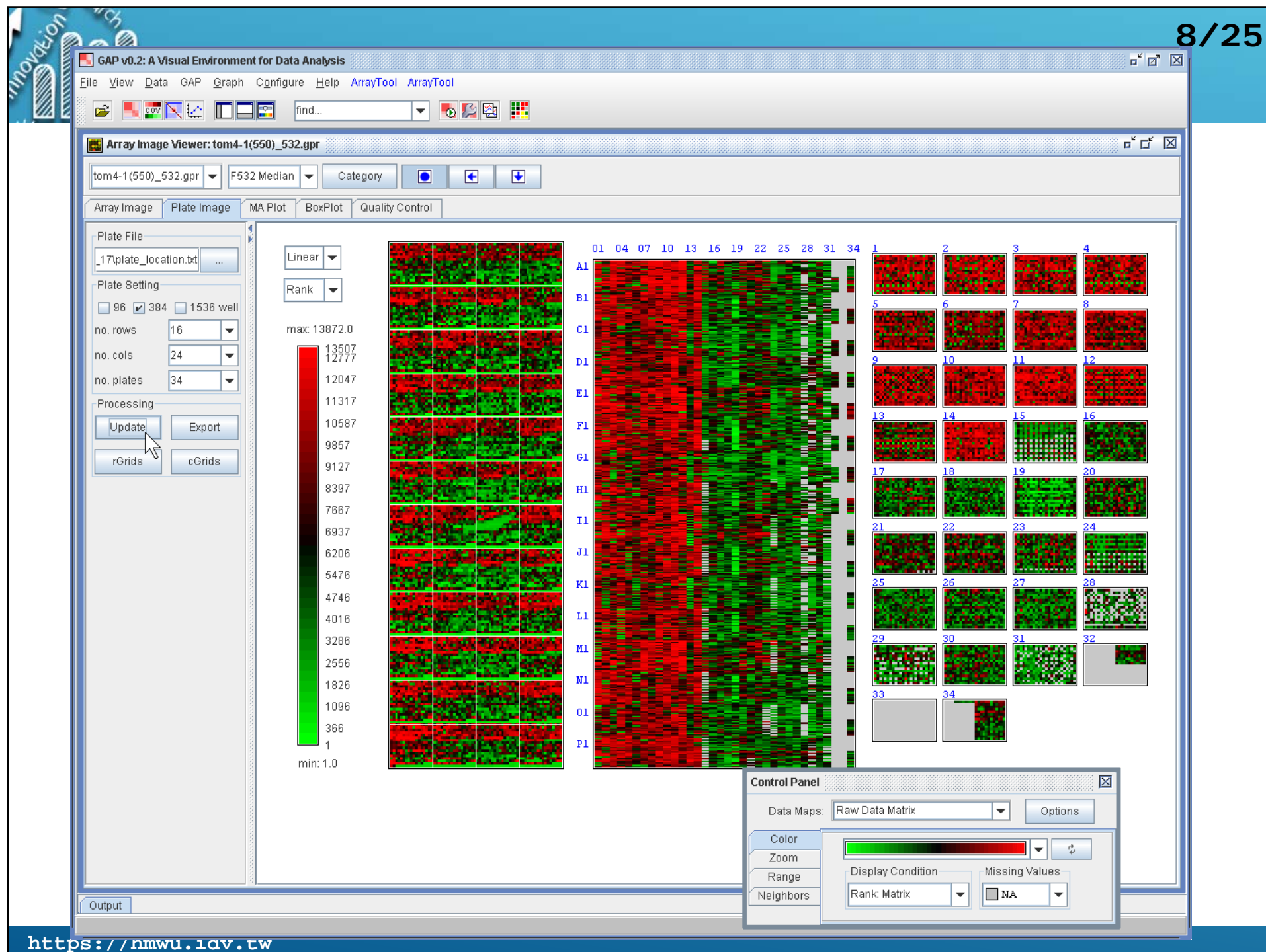
Features:
18 by 18

Signal
16-bit
0~65535

*.gpr

GAL





Applications

D2.01 D2.02 D2.03 D2.04 D2.05 D2.06 D2.07 D2.08 D2.09 D2.10

D7.02 D7.03 D7.04 D7.05 D7.06 D7.07 D7.08 D7.09 D7.10 D7.11

D2.01表示: 2018/06/05, 00:00(含)~01:00共4個時間點(00:00, 00:15, 00:30, 00:45)之停留地區, 取最多停留地區為此時段之停留地區。D7.02表示: 2018/06/10, 01:00(含)~02:00。

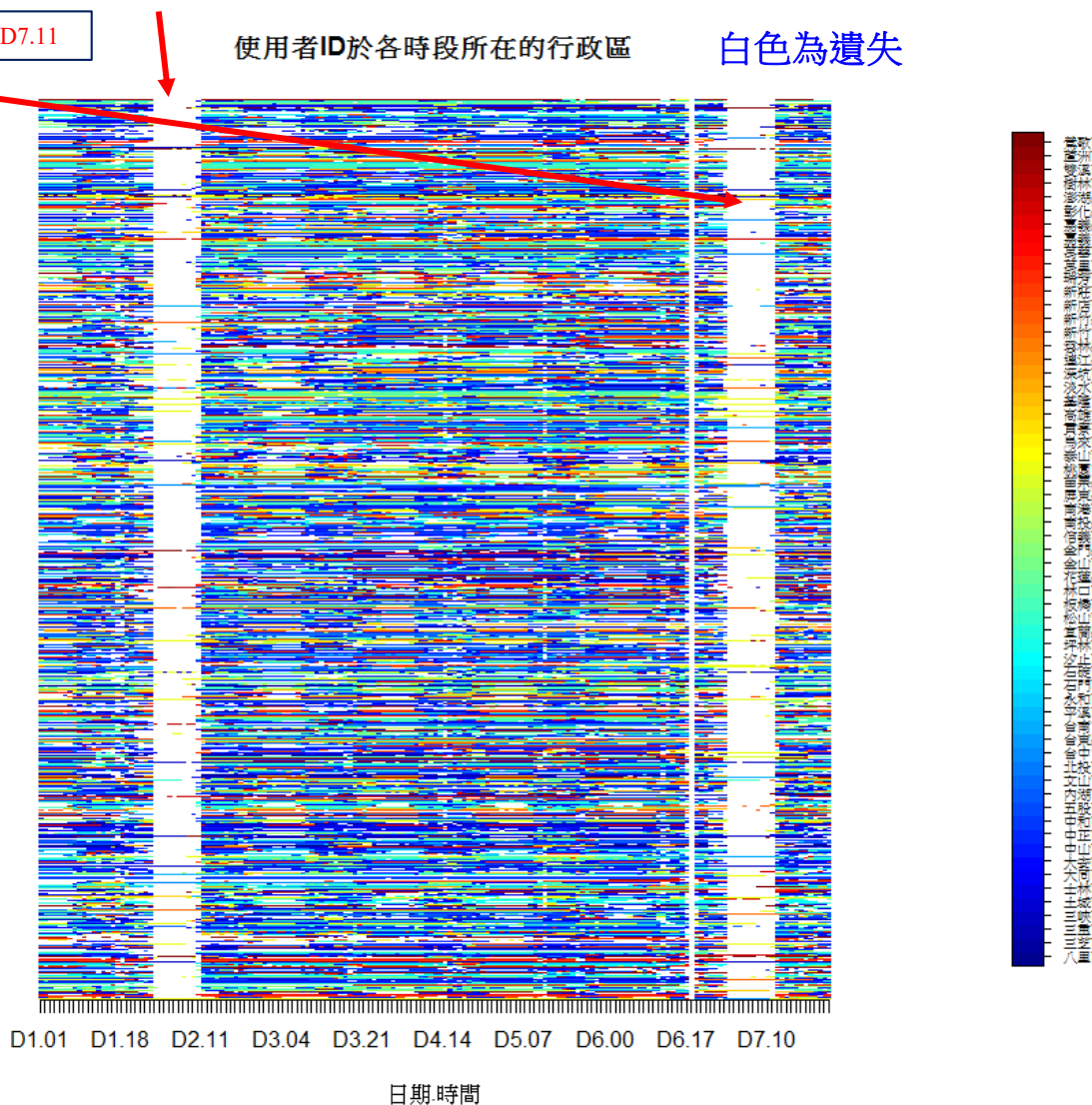
使用者ID於各時段所在的行政區

白色為遺失

id,date,time,latitude,longitude

raw
data

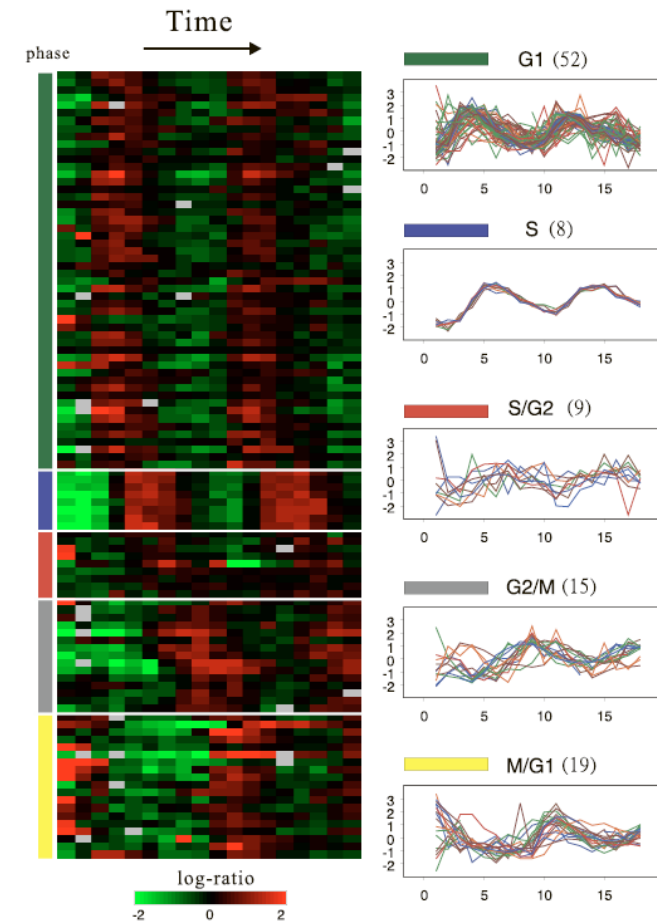
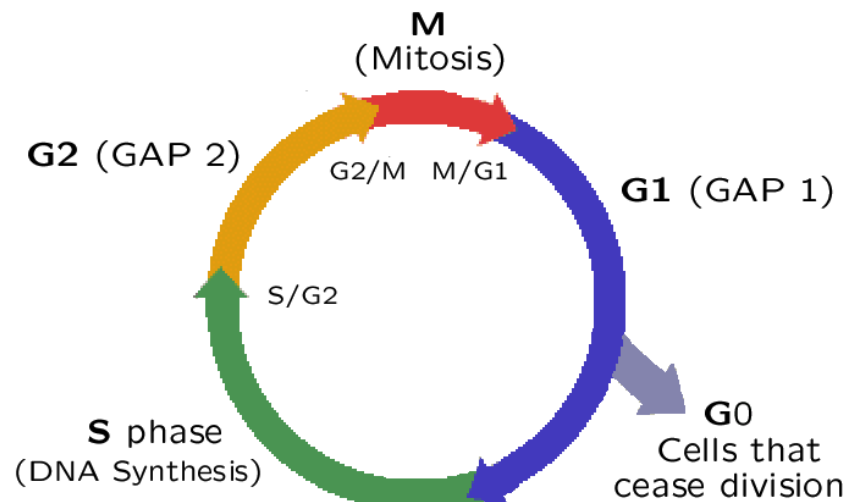

使用者ID



課堂練習: Microarray Data

10/25

- Lu and Wu (2010)
 - Time course data: every 7 minutes and totally 18 time points.
 - Known genes: there are 103 cell cycle-regulated genes by traditional method in G1, S, S/G2, G2/M, or M/G1. (Remove NA's: 79.)



See also: Using R to draw a Heatmap from Microarray Data

http://www2.warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/heatmap/



課堂練習: Microarray Data

11/25

```
# install.packages("fields")
library(fields)
gbr <- two.colors(start="green", middle="black", end="red")
cell.raw <- read.table("trad_alpha103.txt", row.names=1, header=T)
cell.data <- t(scale(t(cell.raw[,2:19]), center=T, scale=T))
n <- nrow(cell.data)
p <- ncol(cell.data)
gene.phase <- cell.raw[,1]
range(cell.data)
cell.data[cell.data > 2.802712] <- 2.802712
cellcycle.color <- c("darkgreen", "blue", "red", "gray50", "orange")
rc <- cellcycle.color[gene.phase+1]
cc <- rainbow(ncol(cell.data))

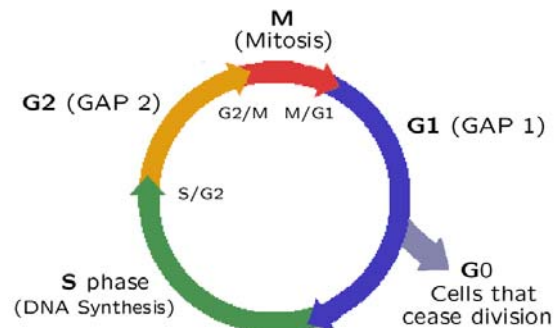
hv1 <- heatmap(cell.data[n:1,], col = gbr, Colv=NA, Rowv=NA,
               RowSideColors = rc,
               ColSideColors = cc, margins = c(5,10),
               xlab = "Times", ylab = "Genes", main = "Heatmap of Microarray Data")

hv2 <- heatmap(cell.data, col = gbr, Colv=NA, Rowv=NULL,
               RowSideColors = rc,
               ColSideColors = cc, margins = c(5,10),
               xlab = "Times", ylab = "Genes", main = "Heatmap of Microarray Data")

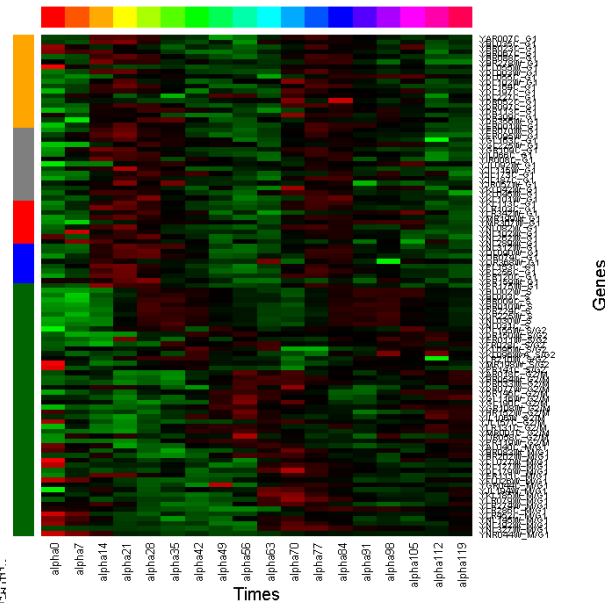
dd <- as.dendrogram(hclust(as.dist(1-cor(t(cell.data)))))
hv3 <- heatmap(cell.data, col = gbr, Colv=NA, Rowv=dd,
               RowSideColors = rc,
               ColSideColors = cc, margins = c(5,10),
               scale = "row",
               xlab = "Times", ylab = "Genes", main = "Heatmap of Microarray Data")
```

課堂練習: Microarray Data

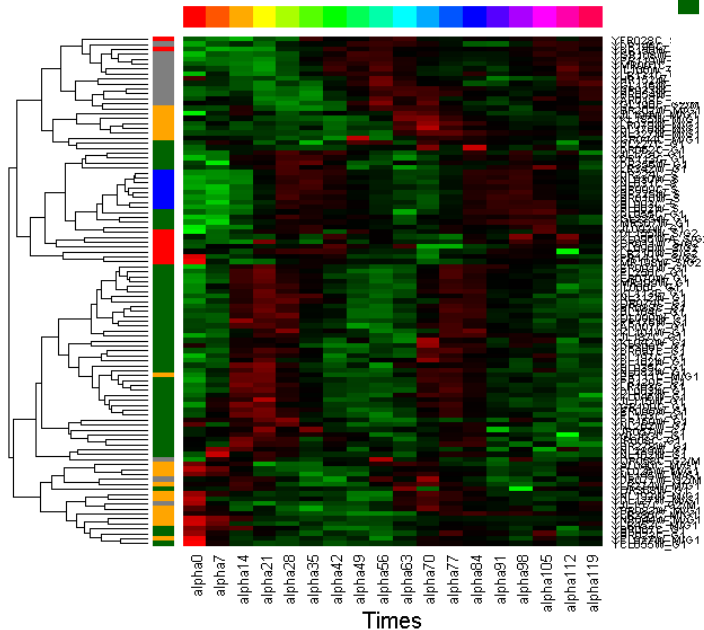
12/25



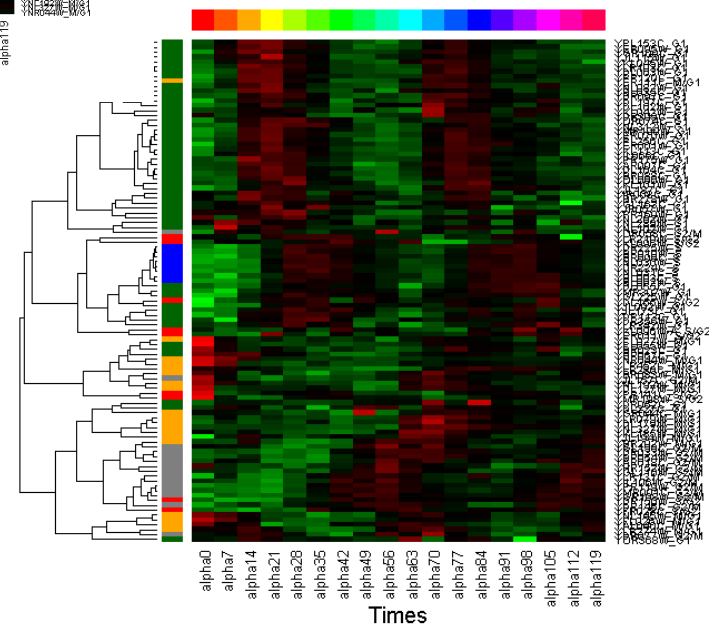
Heatmap of Microarray Data



Heatmap of Microarray Data



Heatmap of Microarray Data





pheatmap: Pretty Heatmaps

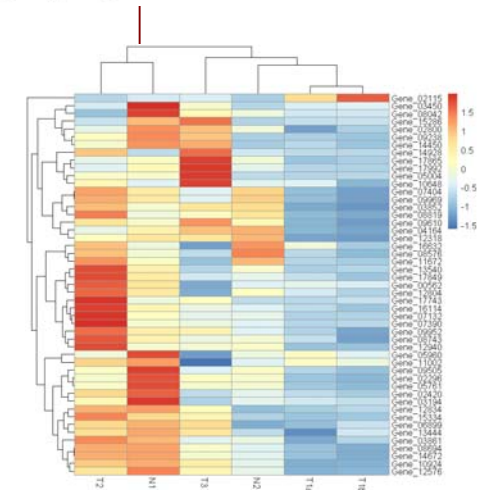
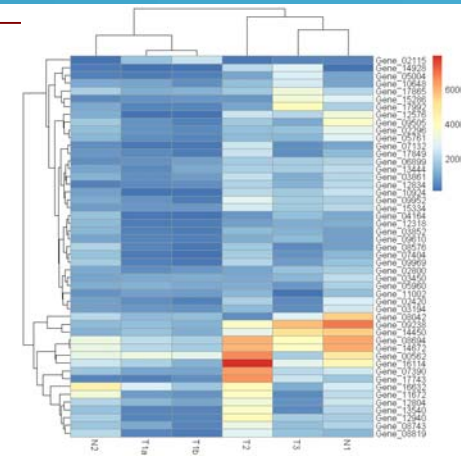
13/25

Implementation of heatmaps that offers more control over dimensions and appearance.

```
> library(pheatmap)
> DESeq_subset <- read.csv("DESeq_subset.csv")
> dim(DESeq_subset)
[1] 49 7
> head(DESeq_subset)
      X  T1a  T1b  T2  T3  N1  N2
1 Gene_00562 32314 29693 66140 17973 47994 30878
2 Gene_02115 15261 23301 1944 4578 4087 1072
...
6 Gene_03194 7611 6806 13506 5727 25020 9235
> DESeq.X <- as.matrix(DESeq_subset[,2:ncol(DESeq_subset)])
> colnames(DESeq.X)
[1] "T1a" "T1b" "T2" "T3" "N1" "N2"
> rownames(DESeq.X) <- DESeq_subset[,1]
> dimnames(DESeq.X)
[[1]]
[1] "Gene_00562" "Gene_02115" "Gene_02296" "Gene_02420" ...
[46] "Gene_17743" "Gene_17849" "Gene_17865" "Gene_17992"

[[2]]
[1] "T1a" "T1b" "T2" "T3" "N1" "N2"

> str(DESeq.X)
int [1:49, 1:6] 32314 15261 6730 ...
- attr(*, "dimnames")=List of 2
 ..$ : chr [1:49] "Gene_00562"...
 ..$ : chr [1:6] "T1a" "T1b" "T2" "T3" ...
> pheatmap(DESeq.X)
```



```
> DESeq.X.std <- t(apply(DESeq.X, 1, scale))
> class(DESeq.X.std)
[1] "matrix"
> dimnames(DESeq.X.std) <- dimnames(DESeq.X)
> pheatmap(DESeq.X.std) # note the color spectrum
```

Making a heatmap in R with the pheatmap package

<https://davetang.org/muse/2018/05/15/making-a-heatmap-in-r-with-the-pheatmap-package/>

<https://hmwu.idv.tw>

pheatmap: Pretty Heatmaps

14/25

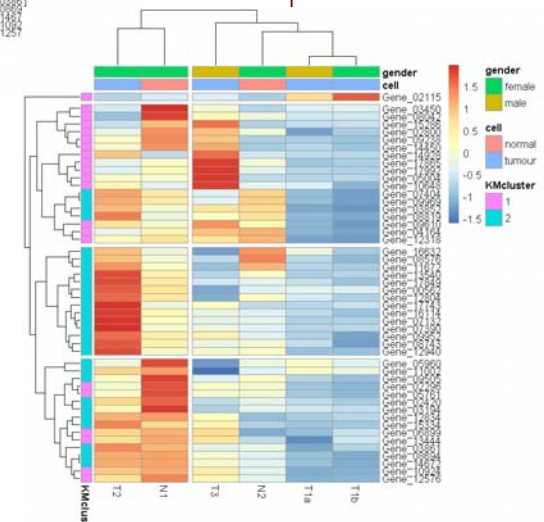
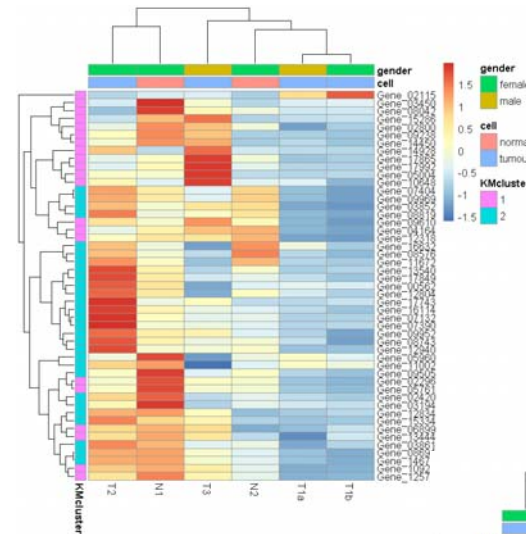
```
> sample.group <- data.frame(cell = rep(c("tumour", "normal"), c(4,2)),
+                               gender = sample(c("male", "female"), 6, replace=T))
> row.names(sample.group) <- colnames(DESeq.X)
> sample.group
```

```
      cell gender
T1a tumour  male
T1b tumour female
T2  tumour female
T3  tumour  male
N1  normal female
N2  normal female
```

```
> km <- as.character(kmeans(DESeq.X.std, 2)$cluster)
> gene.cluster <- data.frame(KMcluster = km)
> row.names(gene.cluster) <- rownames(DESeq.X)
> head(gene.cluster)
```

```
      KMcluster
Gene_00562      2
Gene_02115      1
Gene_02296      1
Gene_02420      2
Gene_02800      1
Gene_03194      2
```

```
>
> pheatmap(DESeq.X.std,
+           annotation_row = gene.cluster,
+           annotation_col = sample.group)
>
> pheatmap(DESeq.X.std,
+           annotation_row = gene.cluster,
+           annotation_col = sample.group,
+           cutree_rows = 4,
+           cutree_cols = 2)
```



display_numbers = TRUE

more examples:

<https://www.jianshu.com/p/86ae39a227f4>

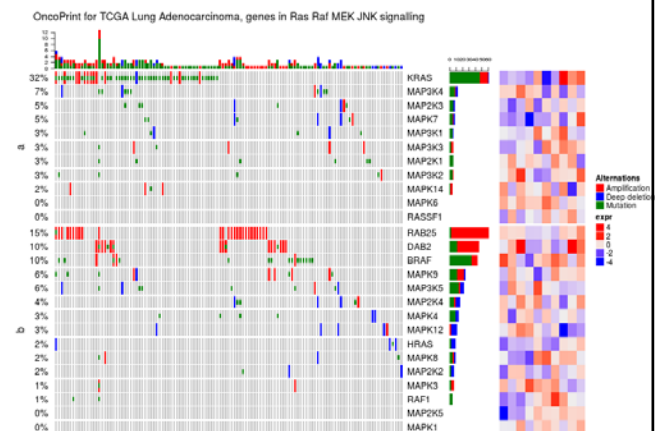
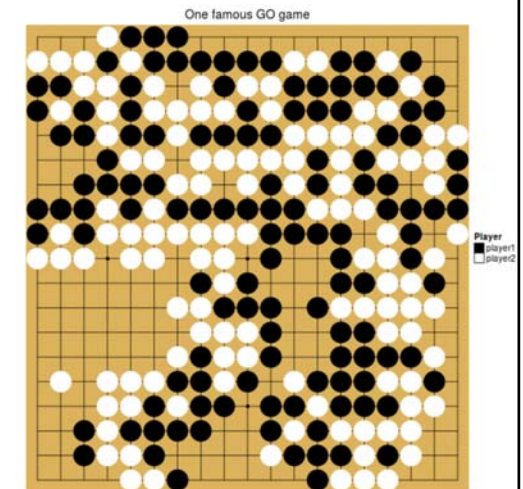
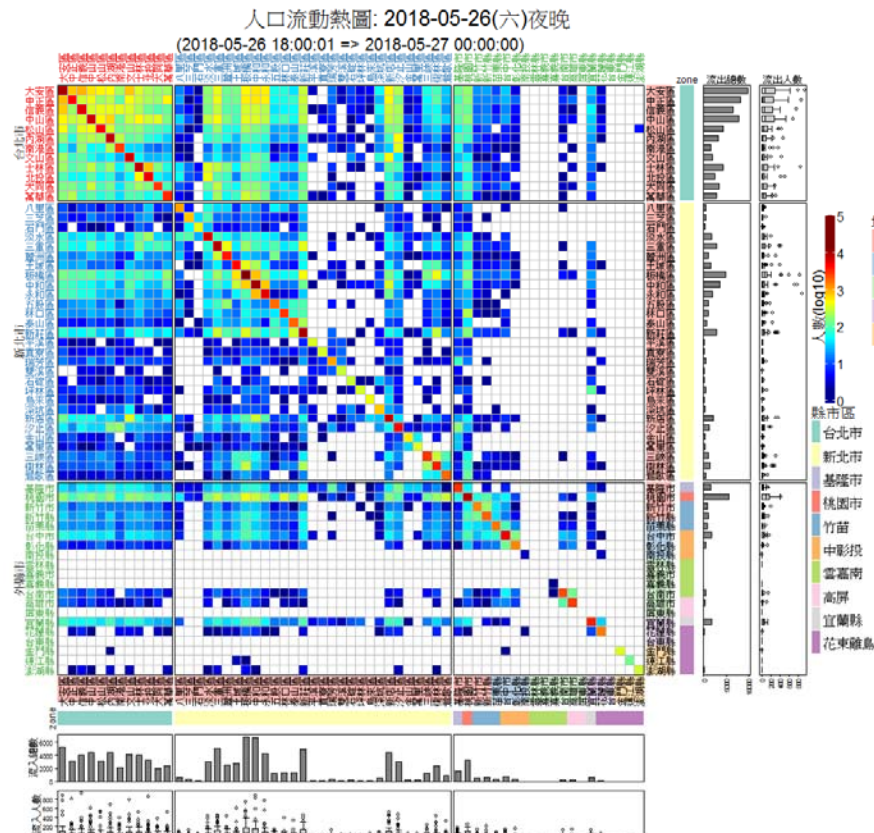
ComplexHeatmap

15/25

<https://jokergoo.github.io/ComplexHeatmap-reference/book/>
<http://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html>

```
> if (!requireNamespace("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
> BiocManager::install("ComplexHeatmap")  
> library(ComplexHeatmap)  
> Heatmap(exprs(selected.eset))
```

Zuguang Gu, Roland Eils, Matthias Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, Bioinformatics, Volume 32, Issue 18, 15 September 2016, Pages 2847–2849.



visualize multiple genomic alteration events by heatmap



Example: ComplexHeatmap

16/25

```
> head(mtcars)

      mpg  cyl  disp  hp drat   wt  qsec vs  am  gear  carb
Mazda RX4    21.0   6  160 110 3.90 2.620 16.46 0   1    4    4
Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 0   1    4    4
Datsun 710    22.8   4  108  93 3.85 2.320 18.61 1   1    4    1
Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44 1   0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02 0   0    3    2
Valiant       18.1   6  225 105 2.76 3.460 20.22 1   0    3    1

> str(mtcars)
'data.frame':   32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...

> mtcars.df <- scale(mtcars)
> class(mtcars.df)
[1] "matrix" "array"
> library(ComplexHeatmap)
> ? Heatmap
```

Reference

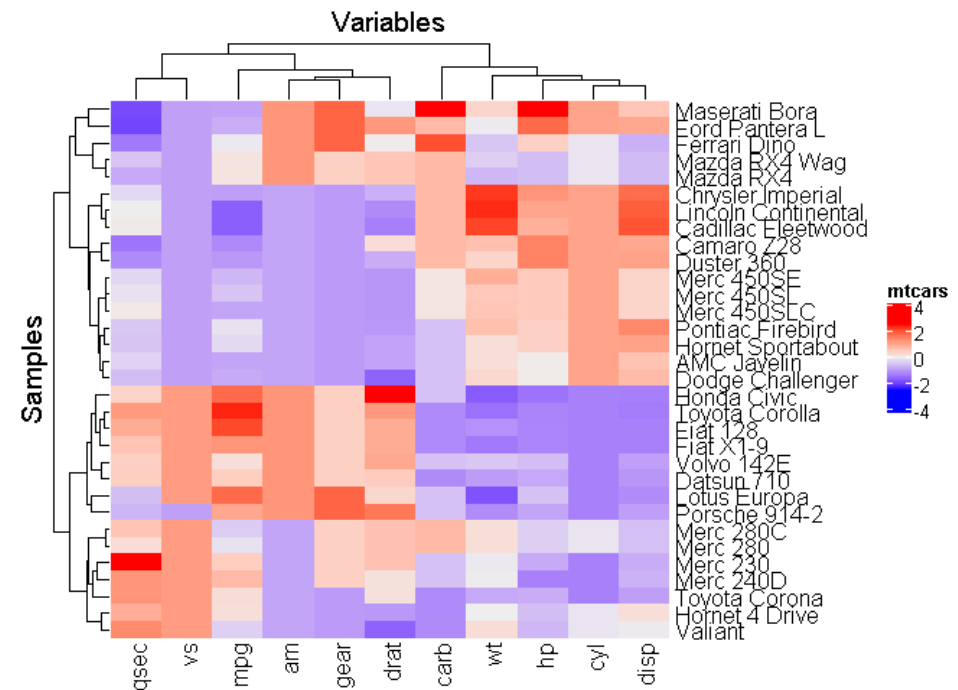
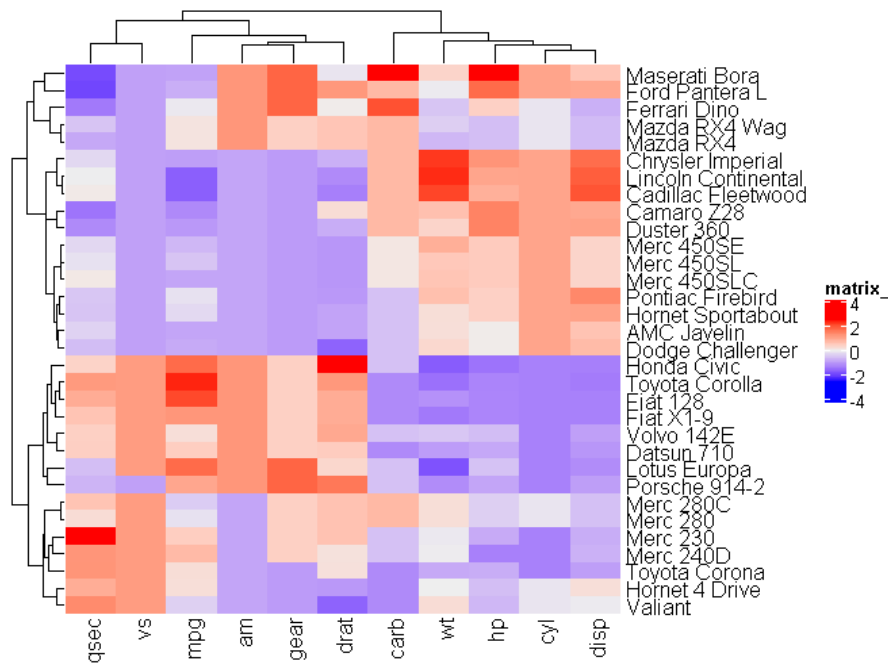
<https://www.datanovia.com/en/lessons/heatmap-in-r-static-and-interactive-visualization/>

ComplexHeatmap: 列、行標題

17/25

```
Heatmap(mtcars.df)
```

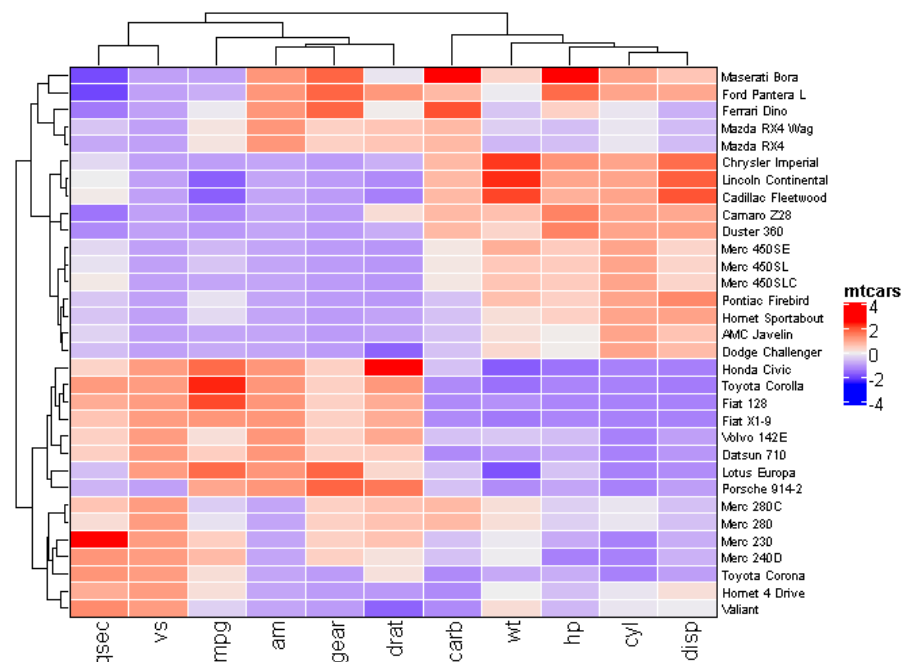
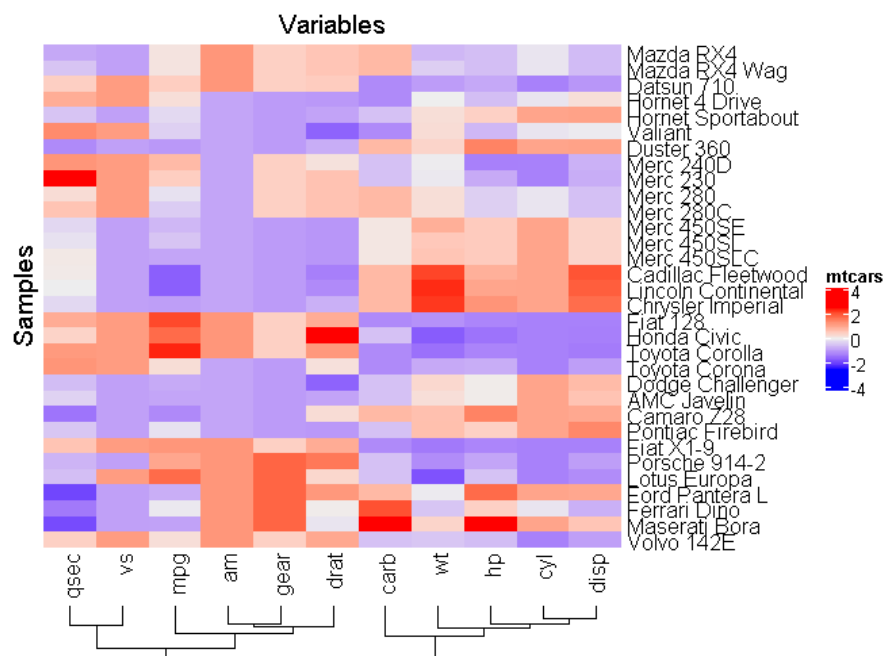
```
Heatmap(mtcars.df, name = "mtcars",  
        row_title = "Samples",  
        column_title = "Variables")
```



樹狀圖位置、格線及字體大小

```
Heatmap(mtcars.df, name = "mtcars",
        row_title = "Samples",
        column_title = "Variables",
        cluster_rows = FALSE,
        column_dend_side = "bottom")
```

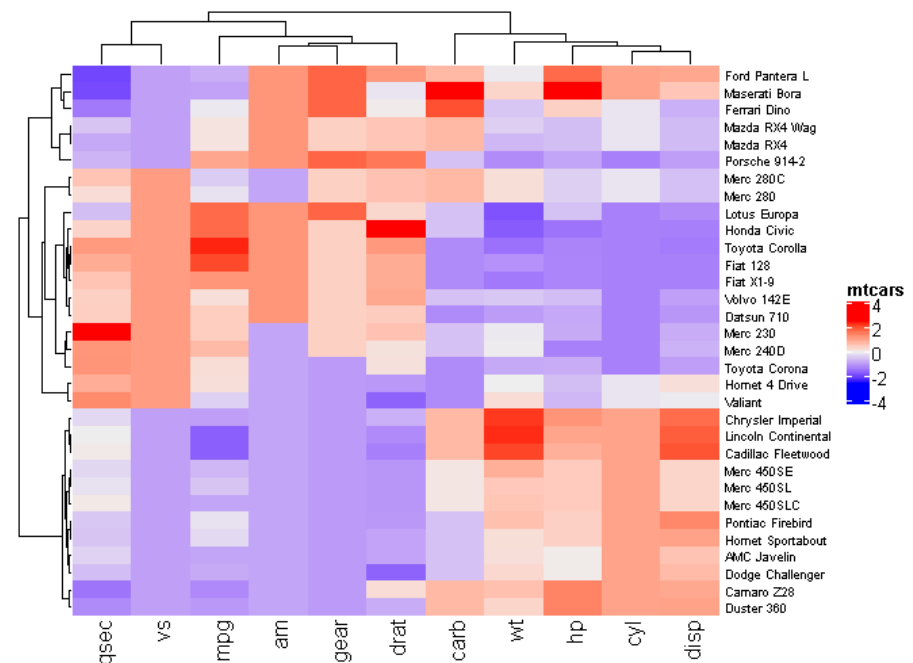
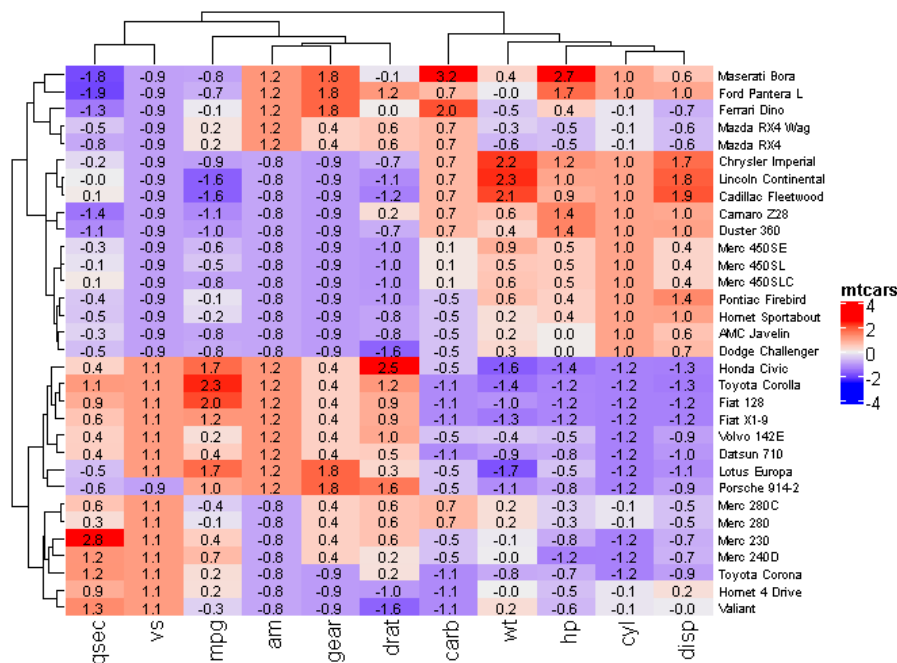
```
Heatmap(mtcars.df, name = "mtcars",
        rect_gp = gpar(col = "white", lwd = 1),
        border = TRUE,
        row_names_gp = gpar(fontsize = 7))
```



ComplexHeatmap: 顯示數字、指定分群法 ^{19/25}

```
Heatmap(mtcars.df, name = "mtcars",  
  cell_fun = function(j, i, x, y, width, height, fill) {  
    grid.text(sprintf("%.1f", mtcars.df[i, j]), x, y,  
      gp = gpar(fontsize = 8))},  
  row_names_gp = gpar(fontsize = 7))
```

```
Heatmap(mtcars.df, name = "mtcars",  
  clustering_distance_rows = "pearson",  
  clustering_method_rows = "average",  
  row_names_gp = gpar(fontsize = 7))
```

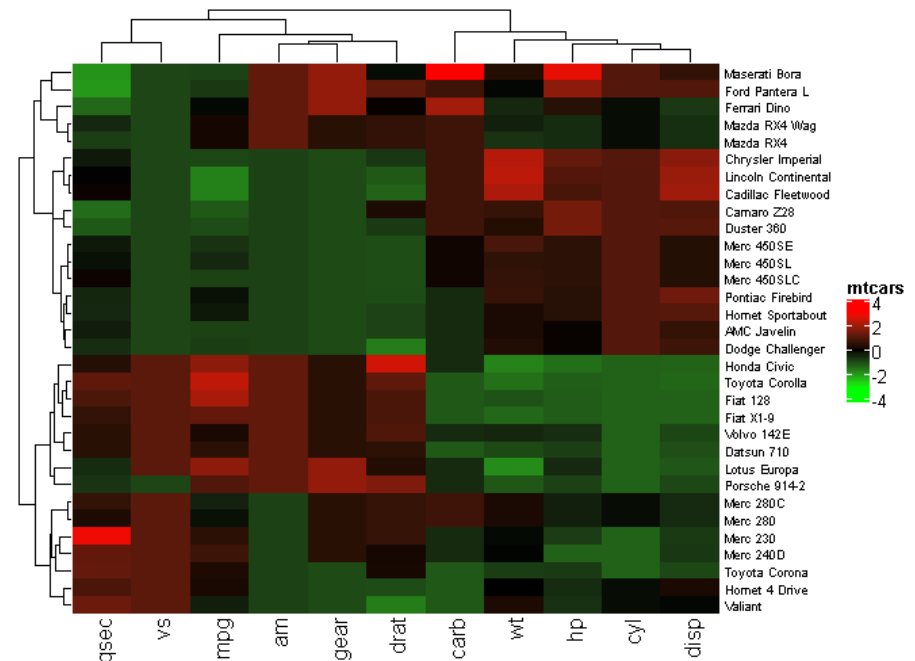
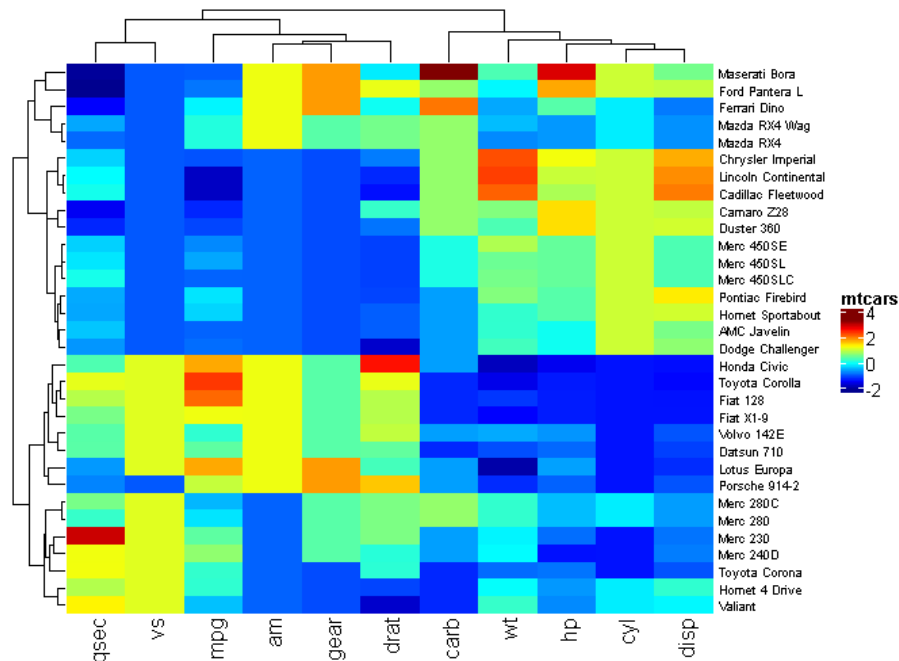


ComplexHeatmap: 自定色階

20/25

```
library(fields)
Heatmap(mtcars.df, name = "mtcars",
        col = tim.colors(),
        row_names_gp = gpar(fontsize = 7))
```

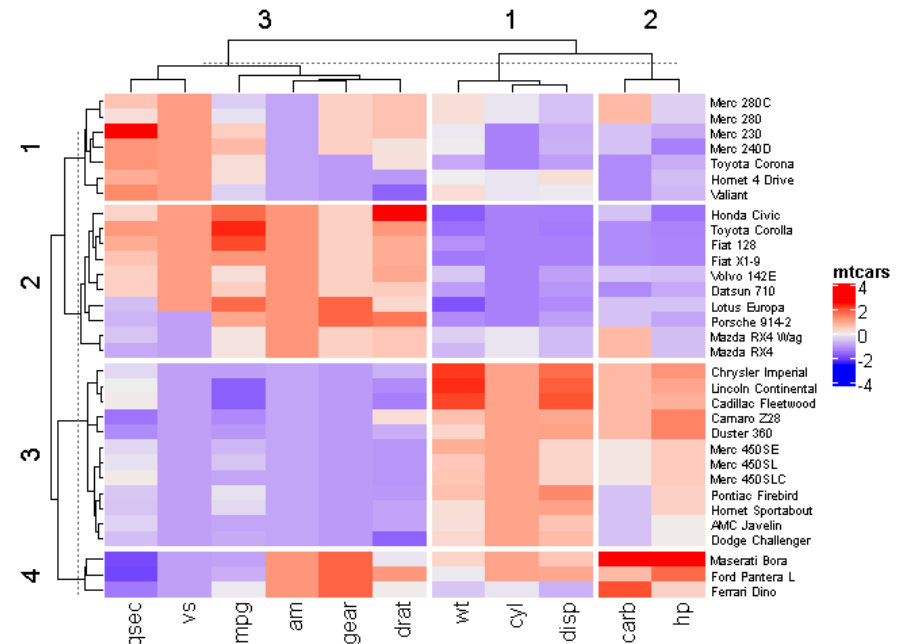
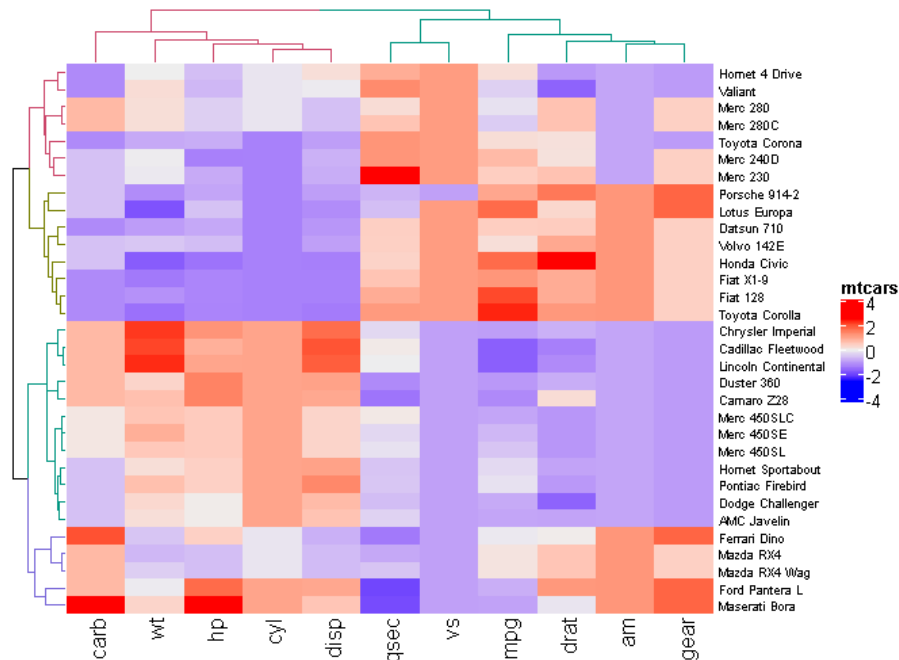
```
library(circlize) # Circular Visualization
mycolorRamp <- colorRamp2(c(-3, 0, 3), c("green", "black", "red"))
Heatmap(mtcars.df, name = "mtcars",
        col = mycolorRamp,
        row_names_gp = gpar(fontsize = 7))
```



以顏色標記樹狀圖、熱圖分割

```
library(dendextend)
row.dend <- hclust(dist(mtcars.df))
column.dend <- hclust(dist(t(mtcars.df)))
Heatmap(mtcars.df, name = "mtcars",
        cluster_rows = color_branches(row.dend, k = 4),
        cluster_columns = color_branches(column.dend, k = 2),
        row_names_gp = gpar(fontsize = 7))
```

```
# split the dendrogram using k-means
Heatmap(mtcars.df, name = "mtcars",
        row_km = 4,
        column_km = 3,
        row_names_gp = gpar(fontsize = 7))
```

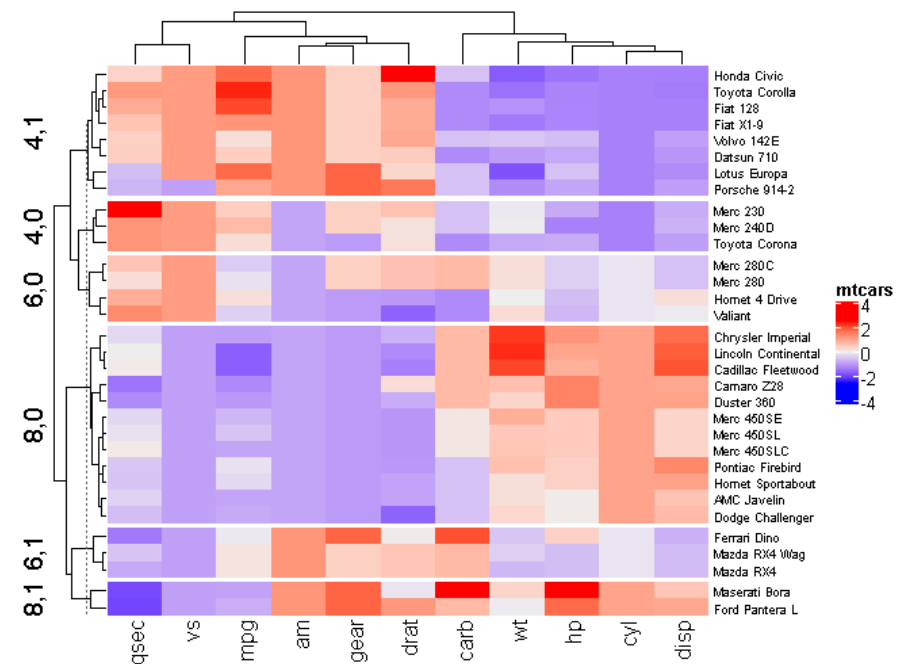
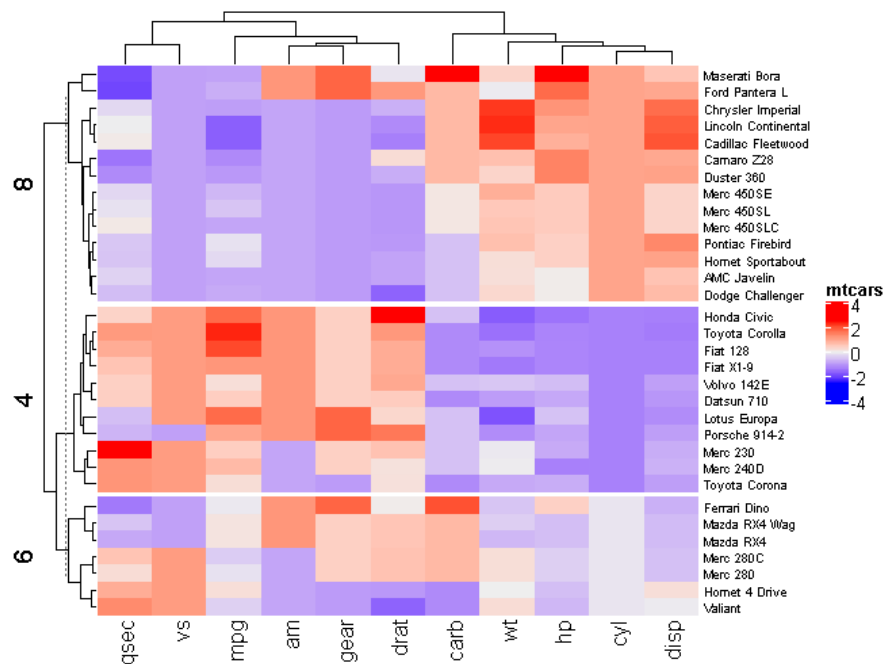


ComplexHeatmap: 依據類別變數分割熱圖

22/25

```
# split by a vector specifying rowgroups
Heatmap(mtcars.df, name = "mtcars",
        row_split = mtcars$cyl,
        row_names_gp = gpar(fontsize = 7))
```

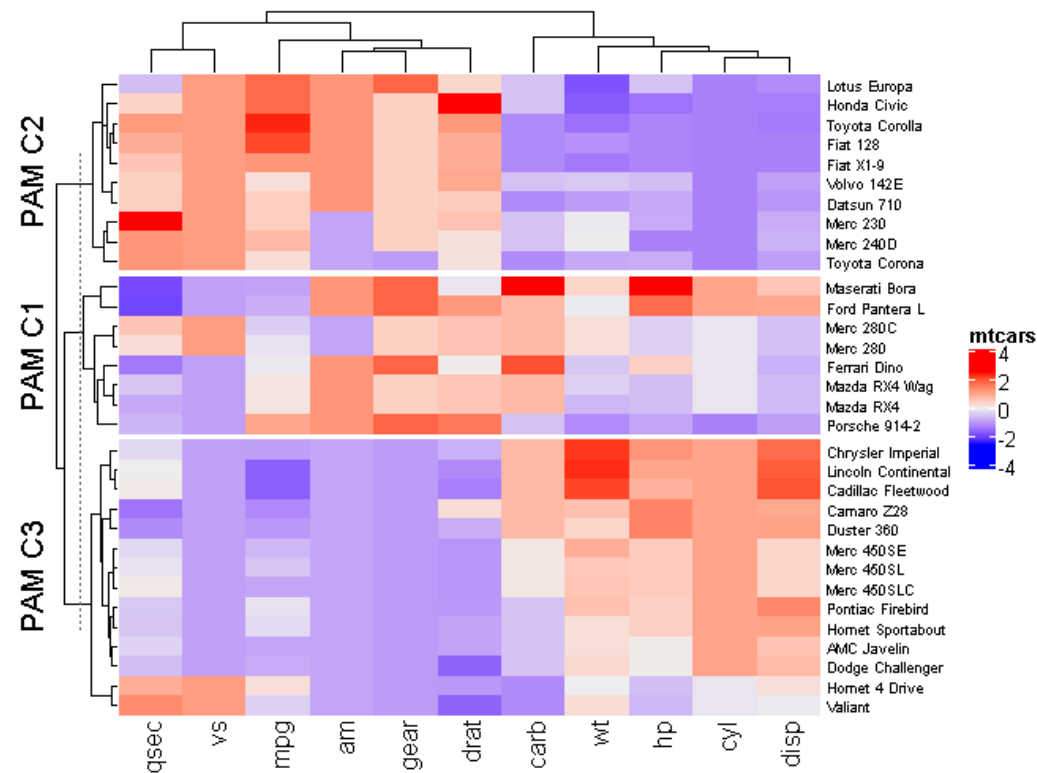
```
# Split by combining multiple variables
Heatmap(mtcars.df, name = "mtcars",
        row_split = data.frame(cyl = mtcars$cyl, am = mtcars$am),
        row_names_gp = gpar(fontsize = 7))
```



ComplexHeatmap: 依據群集分析法分割熱圖

23/25

```
# install.packages("cluster")
library("cluster")
set.seed(12345)
pa <- pam(mtcars.df, k = 3)
Heatmap(mtcars.df, name = "mtcars",
        row_split = paste0("PAM C", pa$clustering),
        row_names_gp = gpar(fontsize = 7))
```

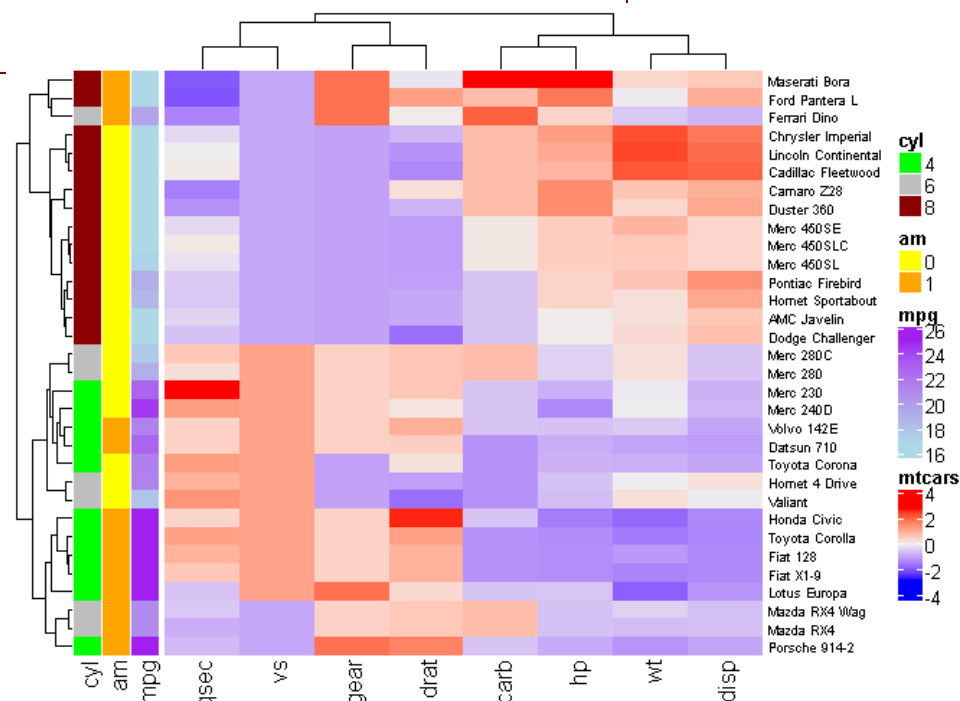


以類別變數標記註解並指定顏色

```
# HeatmapAnnotation for "top_annotation", "bottom_annotation"
# rowAnnotation for "left_annotation", "right_annotation"
var.colors <- list(cyl = c("4" = "green", "6" = "gray", "8" = "darkred"),
                  am = c("0" = "yellow", "1" = "orange"),
                  mpg = colorRamp2(c(17, 25), c("lightblue", "purple")))

ha <- rowAnnotation(cyl = mtcars$cyl, am = mtcars$am, mpg = mtcars$mpg,
                   col = var.colors)

myvars <- colnames(mtcars.df) %in% c("cyl", "am", "mpg")
Heatmap(mtcars.df[, !myvars], name = "mtcars",
       left_annotation = ha,
       row_names_gp = gpar(fontsize = 7))
```





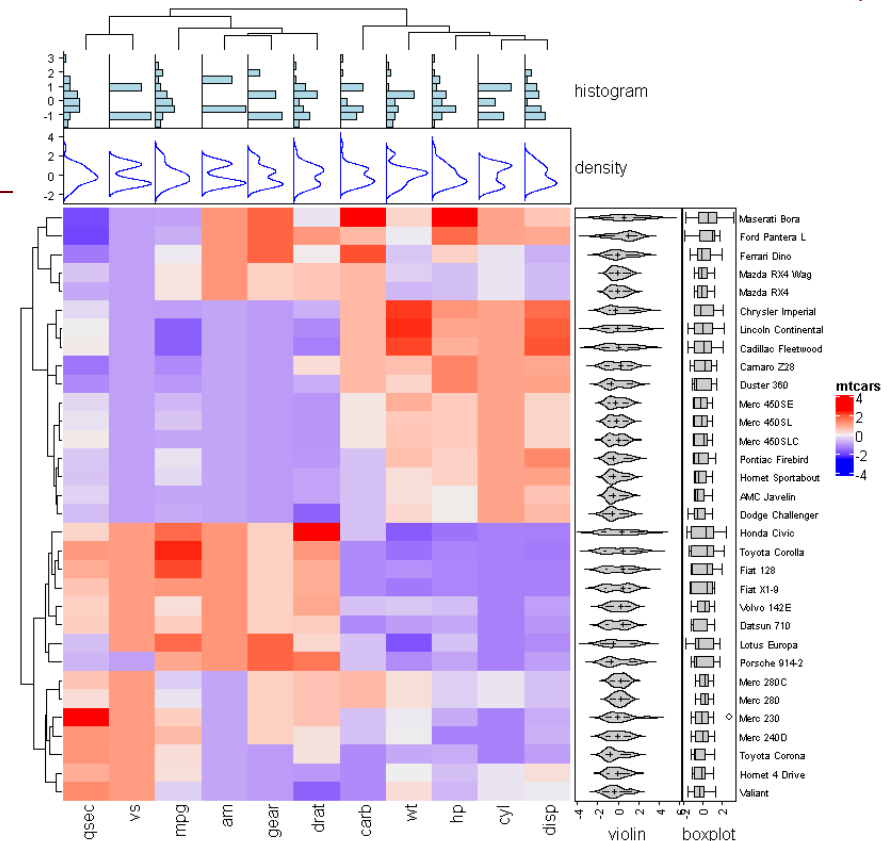
ComplexHeatmap: 標記統計圖為註解

25/25

```
# anno_points, anno_barplot, anno_boxplot, anno_density, anno_histogram
h <- anno_histogram(mtcars.df, gp = gpar(fill = "lightblue"))
d <- anno_density(mtcars.df, type = "line", gp = gpar(col = "blue"))
ha.top <- HeatmapAnnotation(histogram = h, density = d, height = unit(3.8, "cm"))

v <- anno_density(mtcars.df, type = "violin", which = "row")
b <- anno_boxplot(mtcars.df, which = "row")
ha.right <- HeatmapAnnotation(violin = v, boxplot = b, which = "row", width = unit(4, "cm"))

Heatmap(mtcars.df, name = "mtcars",
        top_annotation = ha.top,
        right_annotation = ha.right,
        row_names_gp = gpar(fontsize = 7))
```



Translate from pheatmap to ComplexHeatmap

<https://jokergoo.github.io/2020/05/06/translate-from-pheatmap-to-complexheatmap/>