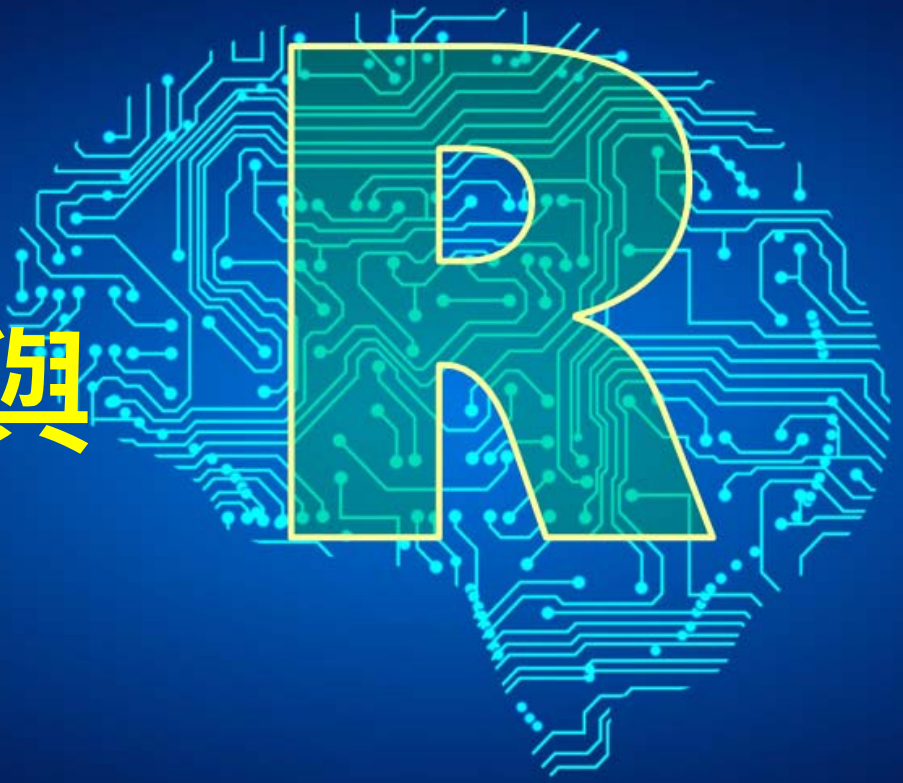
D3-2

台灣人工智慧學校



R統計圖形與資料視覺化

吳漢銘

國立臺北大學 統計學系

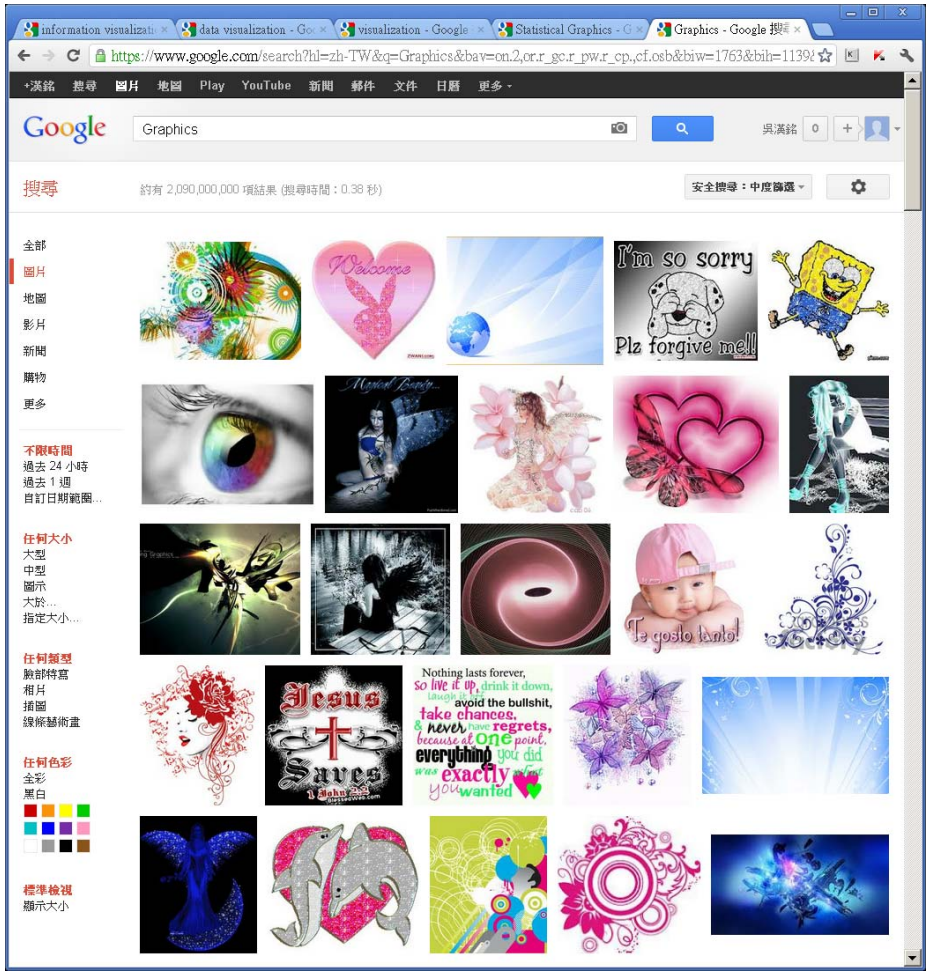
http://www.hmwu.idv.tw

本章大綱

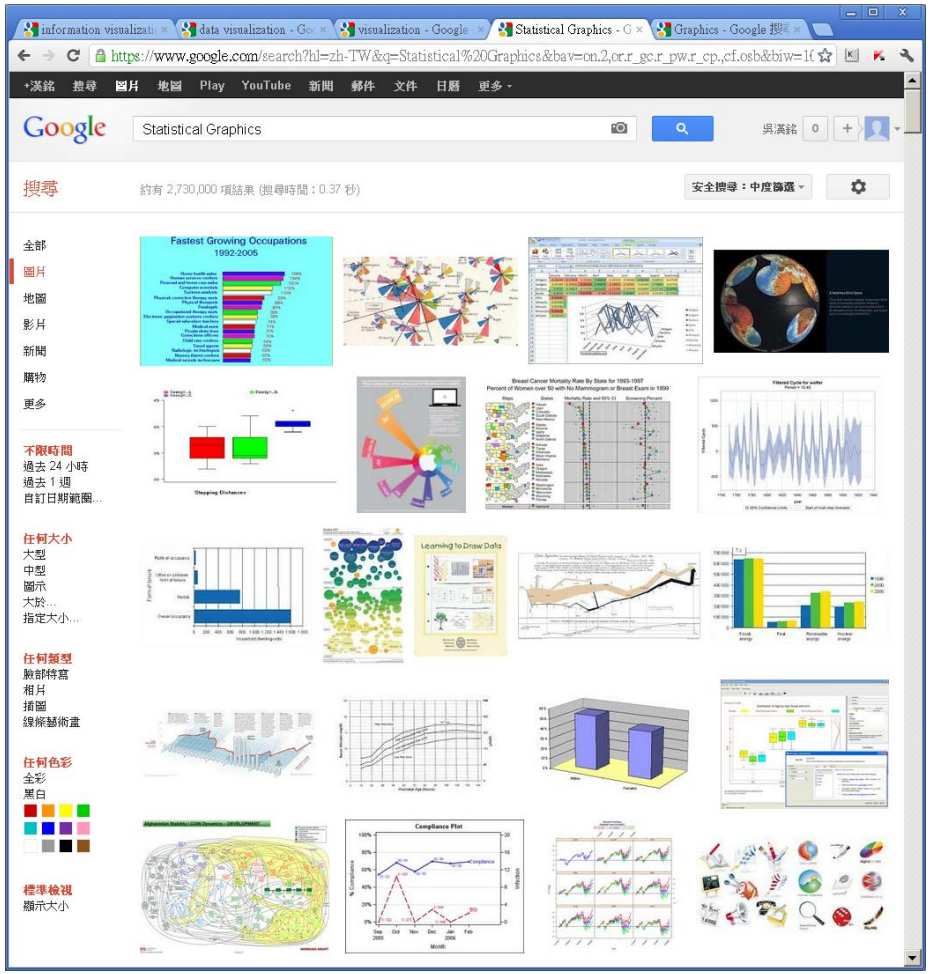
- Data visualization
- **基礎統計圖形**
  - 單一樣本、雙樣本及多樣本的基礎統計圖形
  - 3D散佈圖、rgl、影像、heatmap、等高線圖
- **其它主題**
  - Venn Diagrams、ggplot2、Rgraphviz、igraph、Choropleth Maps、RgoogleMaps、maps, mapdata、googleVis、vcd、ggvis、graphics.SDA
- 巨量資料之視覺化
- 參考書目、Web Resource

Graphics

Graphics

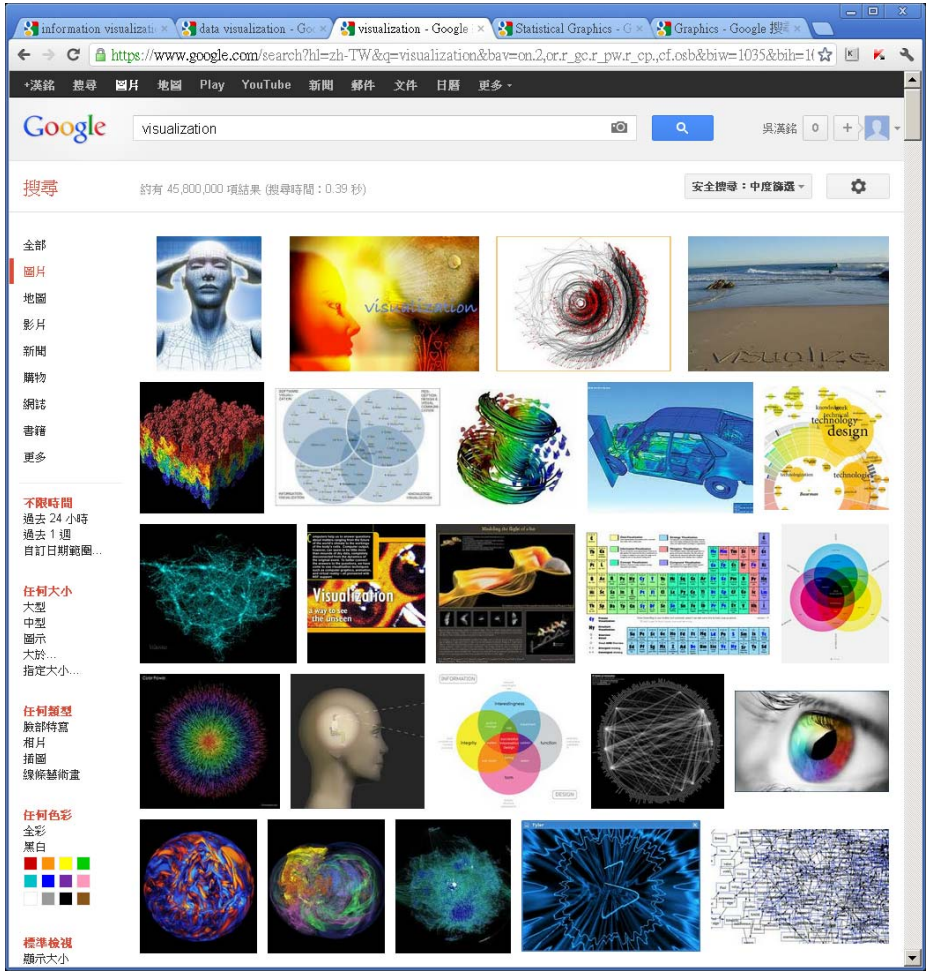


Statistical Graphics

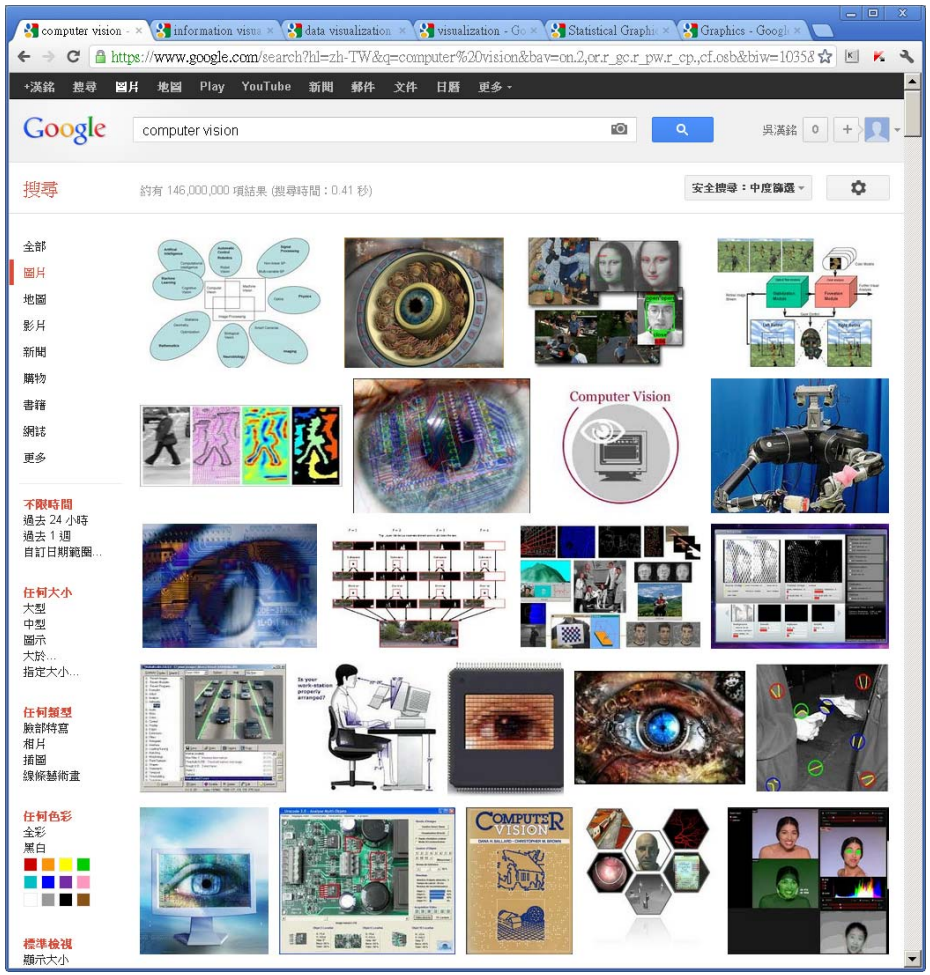


Visualization

Visualization

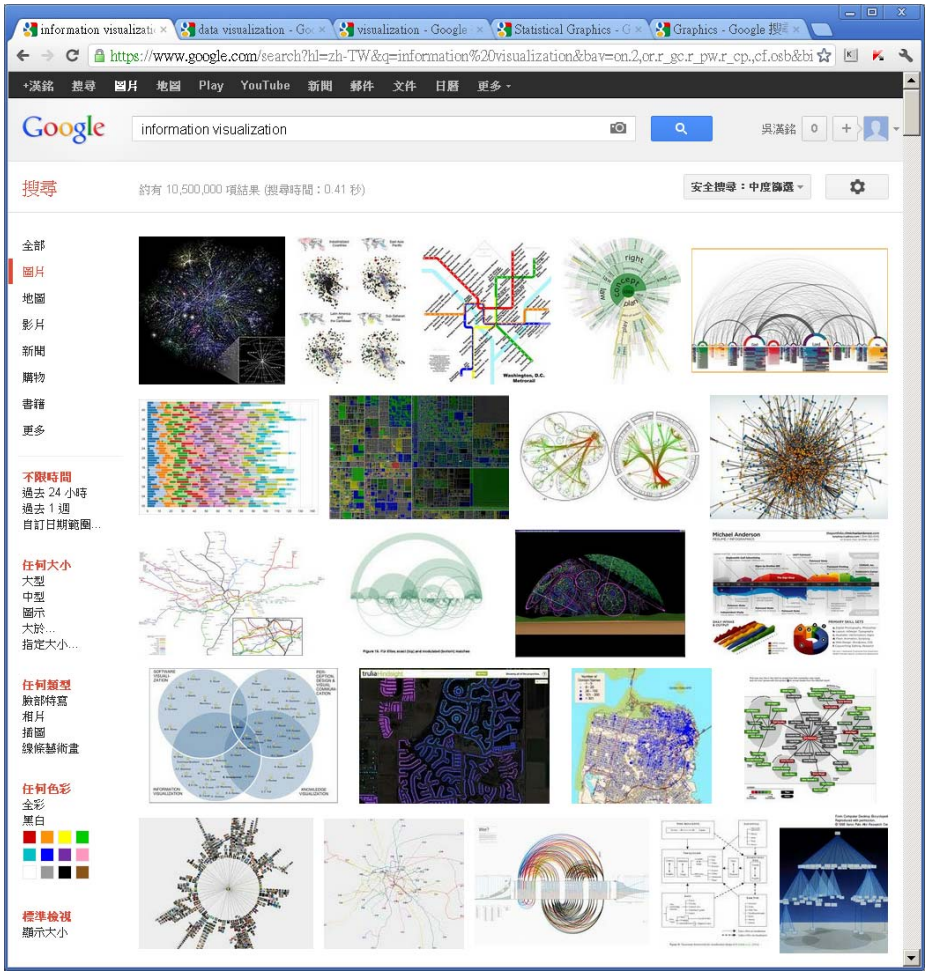


Computer Vision

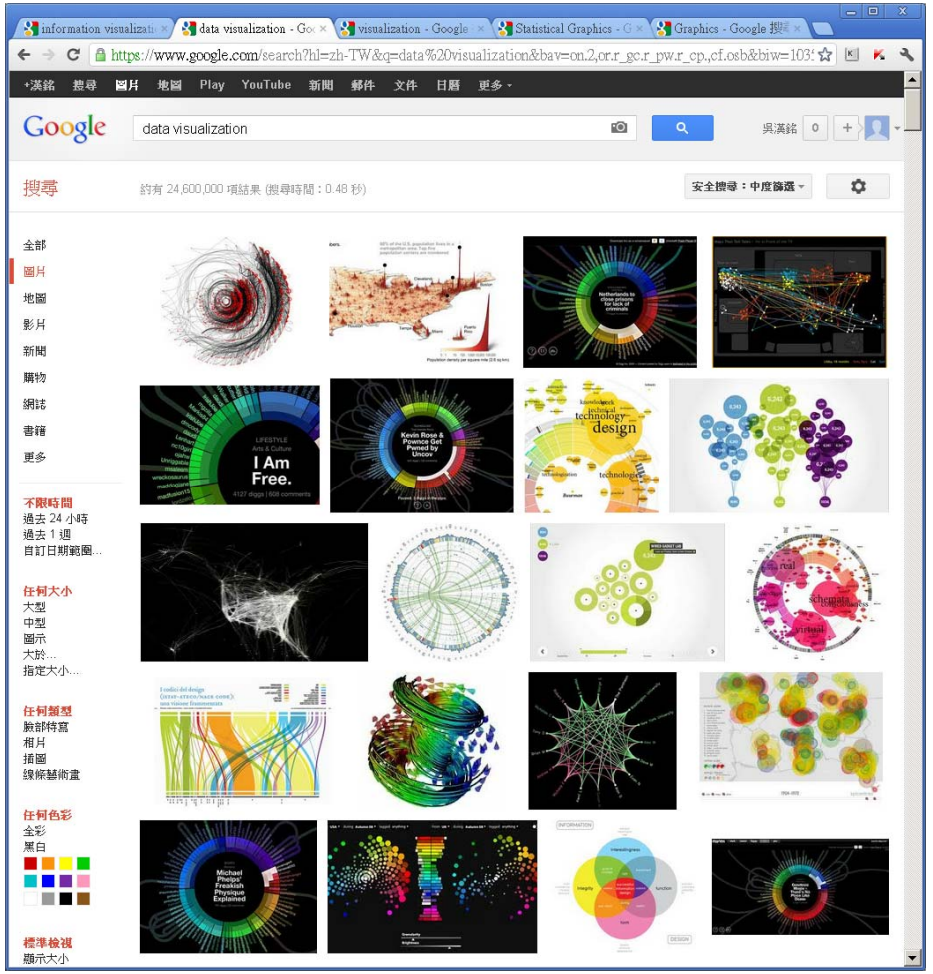


Data Visualization

Information Visualization



Data Visualization



What is Visualization?

People said

- Seeing is believing.
  (眼見為憑)
- Seeing is better than hearing a hundred times.
  (百聞不如一見)
- A picture is worth a thousand words.
  (一幅圖像勝過千言萬語)

The longest name of a city in New Zealand.

The shortest city name in the world is in Norway with one letter (A).

What is visualization?

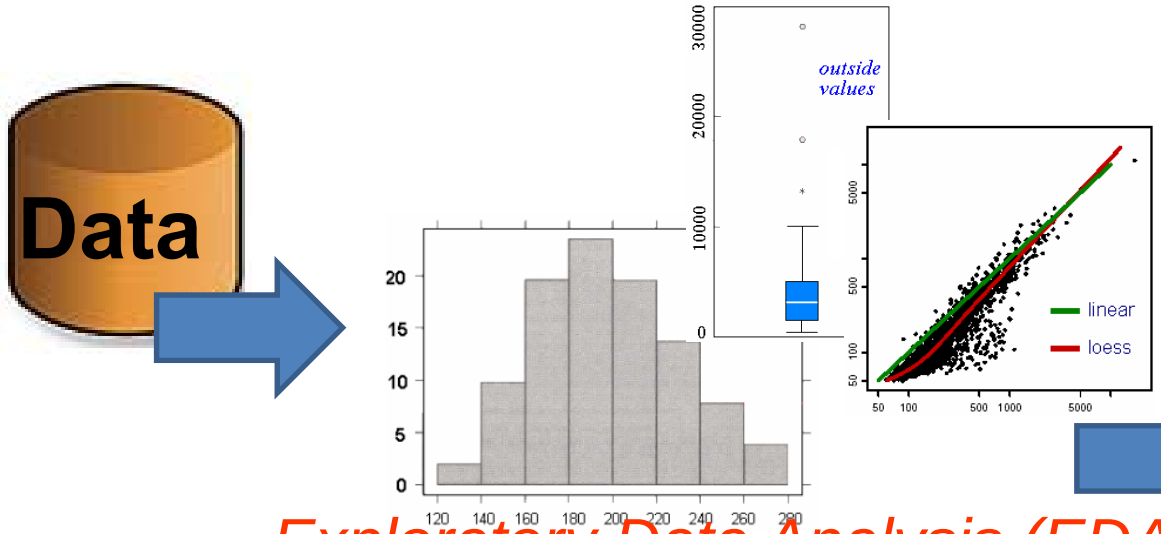
- Making things/processes/abstractions visible (to transform into pictures) that are not directly accessible by the human eye.
- Computer aided extraction and display of information from data.

Picture Source:
http://pictzz.blogspot.com/2009/03/pictures-is-worth-thousand-words.html

Graphical Methods

The purpose of statistical graphics is

to provide visual representations of quantitative information.



information

Exploratory Data Analysis (EDA) Tool

Statistical graphics comprise

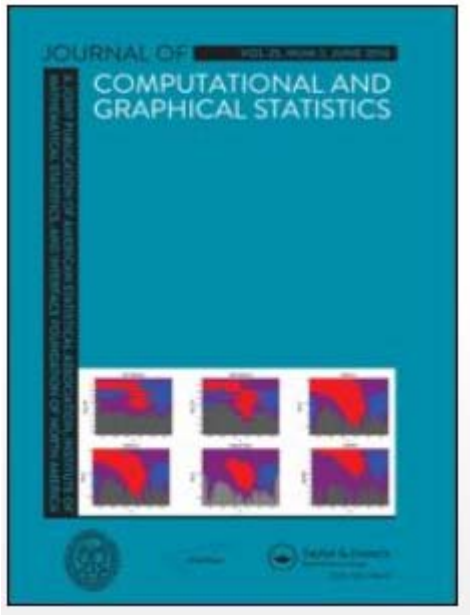
a set of strategies and techniques that provide the research with important insights about the data under examination and help guide the subsequent steps of the research process.

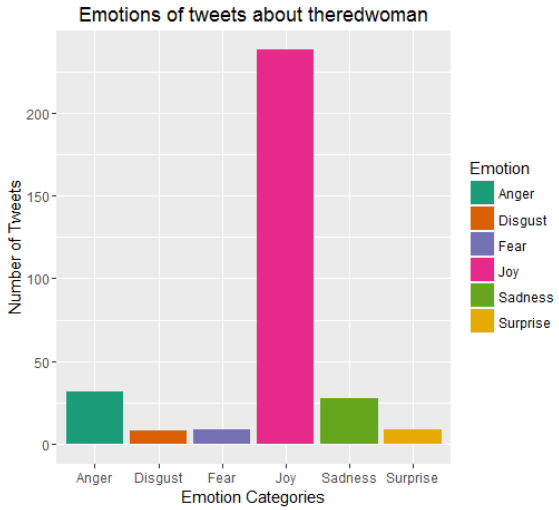
Visualization = Graphing for Data + Fitting + Graphing for Model

Infovis and Statistical Graphics: Different Goals, Different Looks

Journal of Computational and Graphical Statistics, Volume 22, 2013 - Issue 1

- **Infovis and Statistical Graphics: Different Goals, Different Looks**
  Andrew Gelman & Antony Unwin, Pages: 2-28
- **InfoVis Is So Much More**: A Comment on Gelman and Unwin and an Invitation to Consider the Opportunities, Robert Kosara, Pages: 29-32
- **InfoVis and Statistical Graphics: Comment**
  Paul Murrell, Pages: 33-37
- **Graphical Criticism: Some Historical Notes**
  Hadley Wickham , Pages: 38-44
- **Tradeoffs in Information Graphics**
  Andrew Gelman & Antony Unwin , Pages: 45-49





http://emarketingwall.com/how-twitter-responded-to-the-latest-episode-of-game-of-thrones

Why Data Visualization?

- It is not about "**infographics**", the beautiful, heavily customized products of expert graphic designers.
- Data visualization can provide clear understanding of patterns in data, detect hidden structures in data, condense information.
- **Anscombe's quartet** comprises four datasets. They were constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing data before analyzing it and the effect of outliers on statistical properties.
- Four datasets have nearly identical simple statistical properties, yet appear very different when graphed.

	I		II		III		IV	
	x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.1	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.1	4	5.39	19	12.5
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89

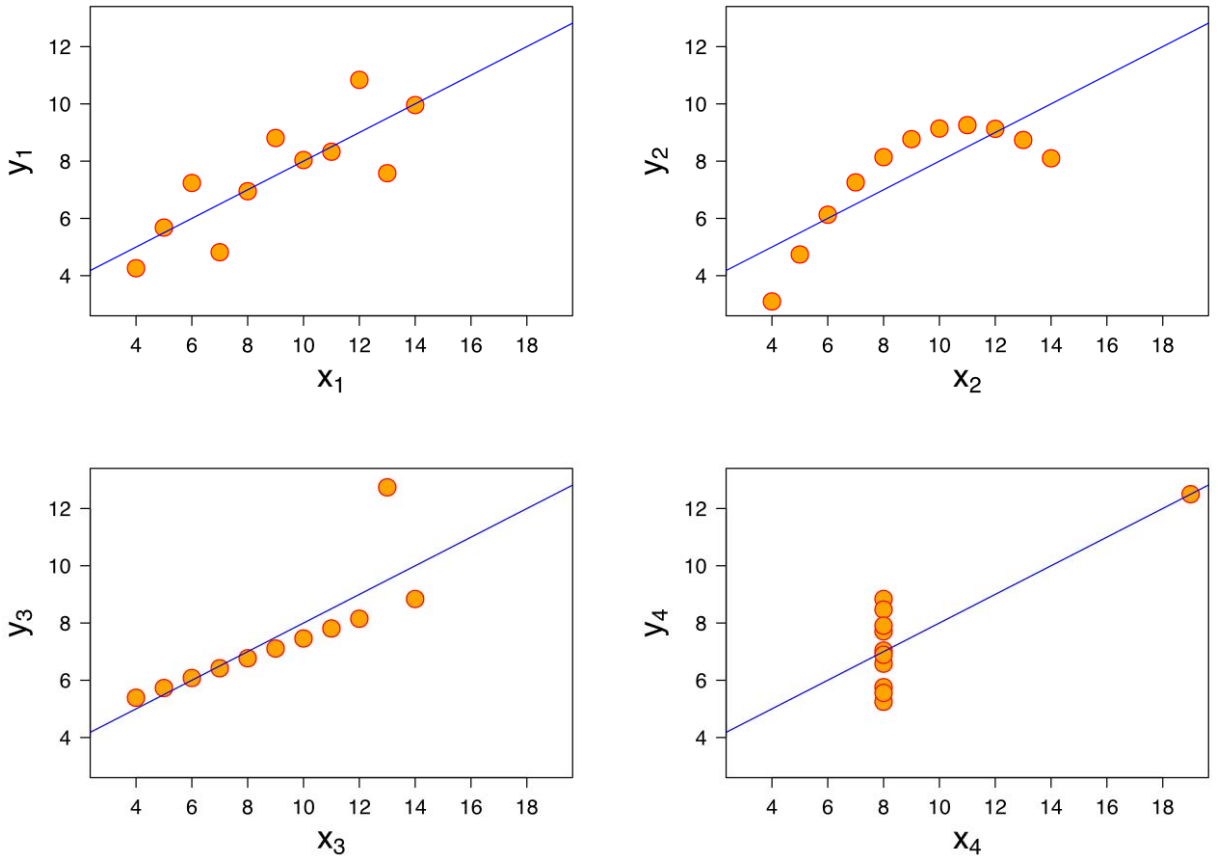
Mean of x in each case: 9 (exact)
Sample variance of x in each case: 11 (exact)
Mean of y in each case: 7.50 (to 2 decimal places)
Sample variance of y in each case: 4.122 or 4.127 (to 3 decimal places)
Correlation between x and y in each case: 0.816 (to 3 decimal places)
Linear regression line in each case: y = 3.00 + 0.500x (to 2 and 3 decimal places, respectively)

https://en.wikipedia.org/wiki/Anscombe's_quartet

http://ryanwomack.com/IASSIST/DataViz/

Anscombe's Quartet

- **Mean of x** in each case: 9 (exact)
- **Sample variance of x** in each case: 11 (exact)
- **Mean of y** in each case: 7.50 (to 2 decimal places)
- **Sample variance of y** in each case: 4.122 or 4.127 (to 3 decimal places)
- **Correlation** between x and y in each case: 0.816 (to 3 decimal places)
- **Linear regression line** in each case: y = 3.00 + 0.500x (to 2 and 3 decimal places, respectively)



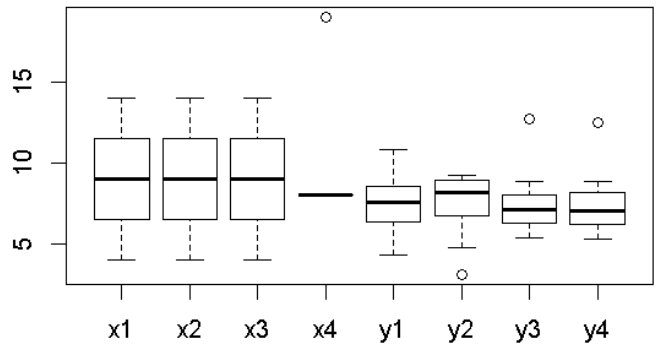
anscombe {datasets}

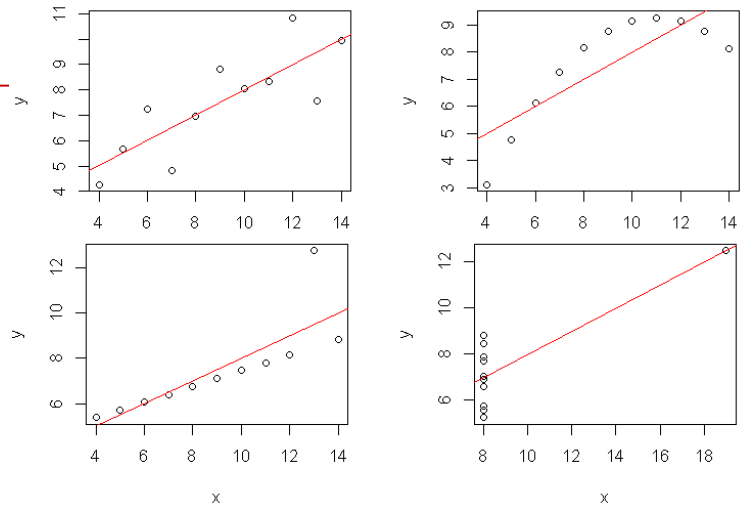
Anscombe's Quartet of 'Identical' Simple Linear Regressions

```
> head(anscombe, 3)
   x1 x2 x3 x4    y1   y2    y3    y4
1  10 10 10  8  8.04 9.14  7.46  6.58
2   8  8  8  8  6.95 8.14  6.77  5.76
3  13 13 13  8  7.58 8.74 12.74  7.71
> apply(anscombe, 2, mean)
      x1       x2       x3       x4       y1       y2       y3       y4
9.000000 9.000000 9.000000 9.000000 7.500909 7.500909 7.500000 7.500909
> apply(anscombe, 2, sd)
      x1       x2       x3       x4       y1       y2       y3       y4
3.316625 3.316625 3.316625 3.316625 2.031568 2.031657 2.030424 2.030579
> mapply(cor, anscombe[,1:4], anscombe[,5:8])
       x1        x2        x3        x4
0.8164205 0.8162365 0.8162867 0.8165214
> mapply(function(x, y) lm(y~x)$coefficients, anscombe[, 1:4], anscombe[, 5:8])
                 x1       x2        x3        x4
(Intercept) 3.0000909 3.000909 3.0024545 3.0017273
x           0.5000909 0.500000 0.4997273 0.4999091
```

```
par(mfrow=c(2, 2))
regplot <- function(x, y){
  plot(y~x)
  abline(lm(y~x), col="red")
}
mapply(regplot, anscombe[, 1:4], anscombe[, 5:8])
```

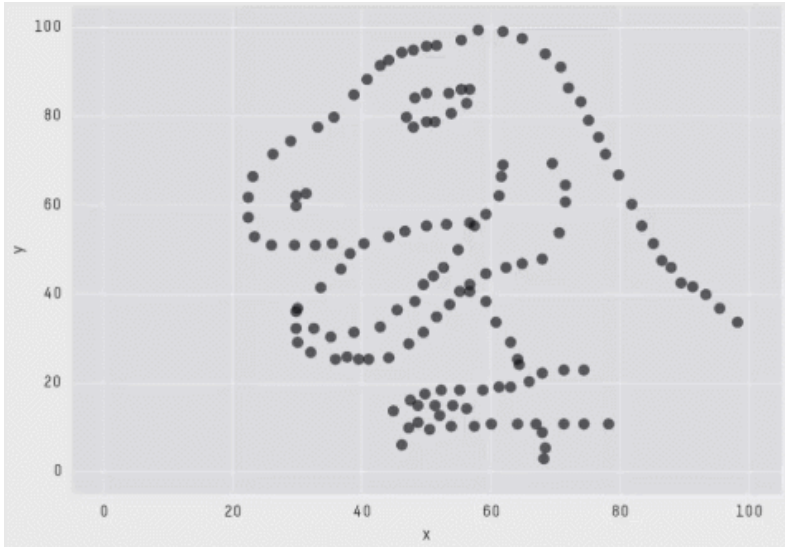
boxplot(anscombe)





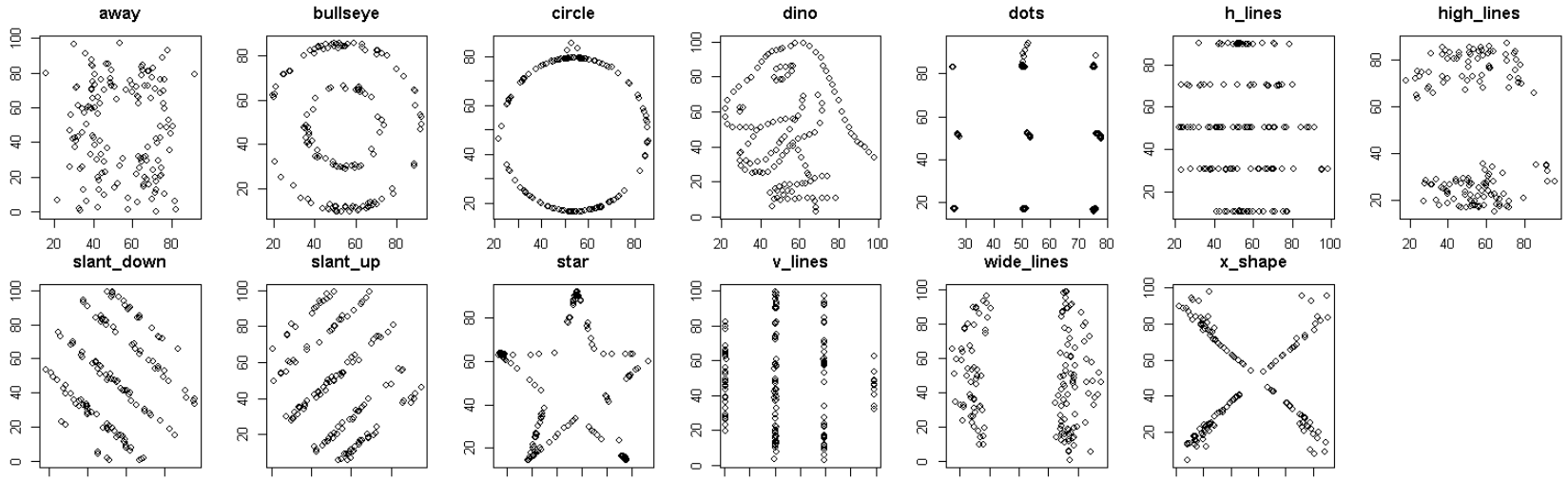
The Datasaurus Dozen

```
install.packages("datasauRus")
```

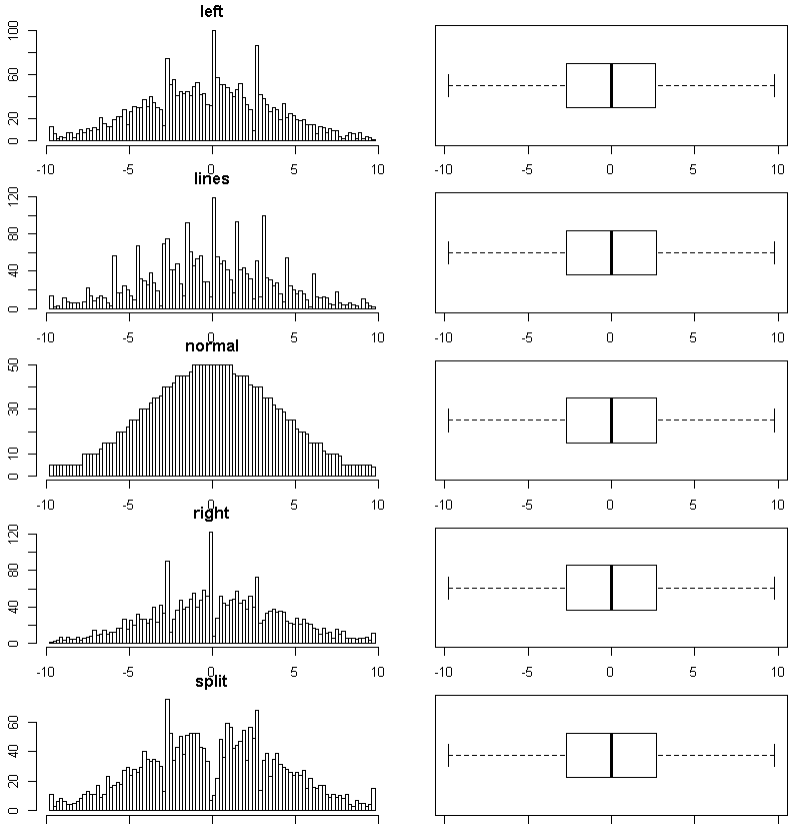


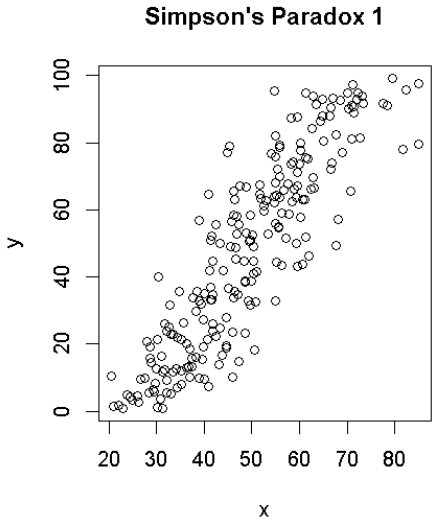

```
X Mean:  54.26 59224
Y Mean:  47.83 13999
X SD   :  16.76 49829
Y SD   :  26.93 42120
Corr.  :  -0.06 42526
```

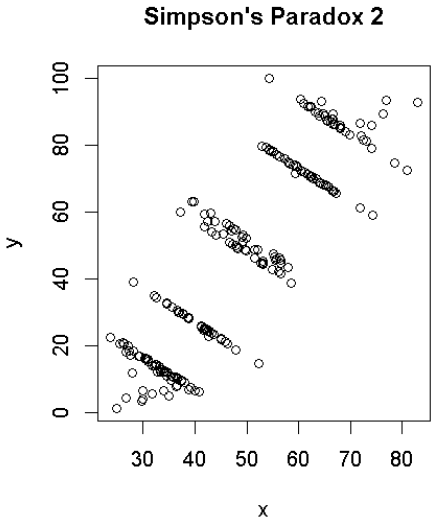
Justin Matejka and George Fitzmaurice, Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. https://www.autodeskresearch.com/publications/samestats



The Datasaurus Dozen
More examples





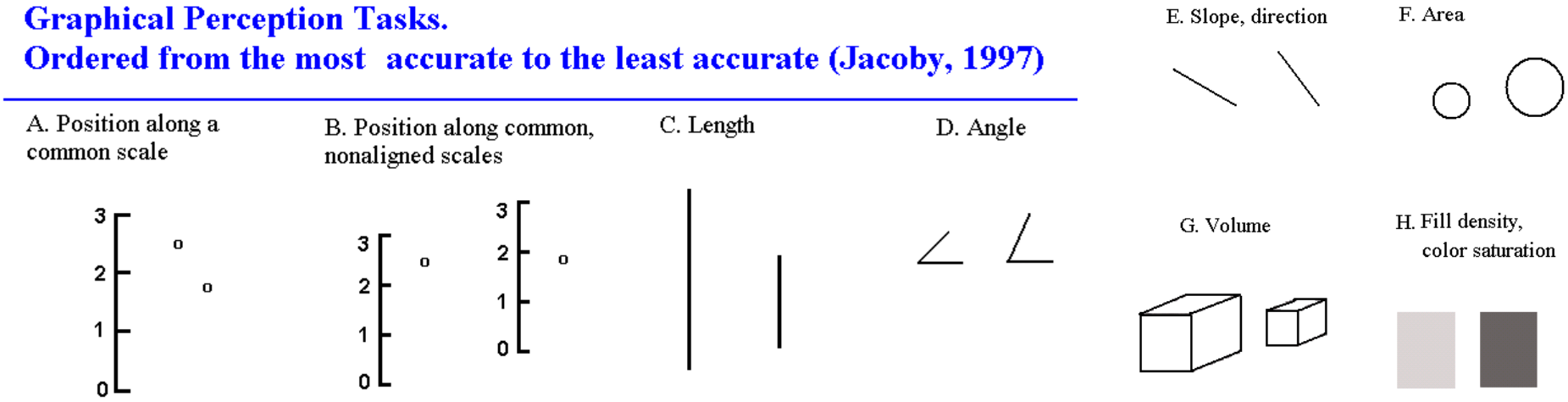


Graphical Perception

Human reception and comprehension of graphical information involves three fundamental perceptual task:

- Detection: the visual recognition of a geometric aspect that encodes a physical value. The basic information from the data must be discernible in the graph.
- Assembly: the process of discerning patterned regularities among the discrete elements of a graphical display.
- Estimation: the visual assessment of the relative magnitudes of two or more quantitative physical values.

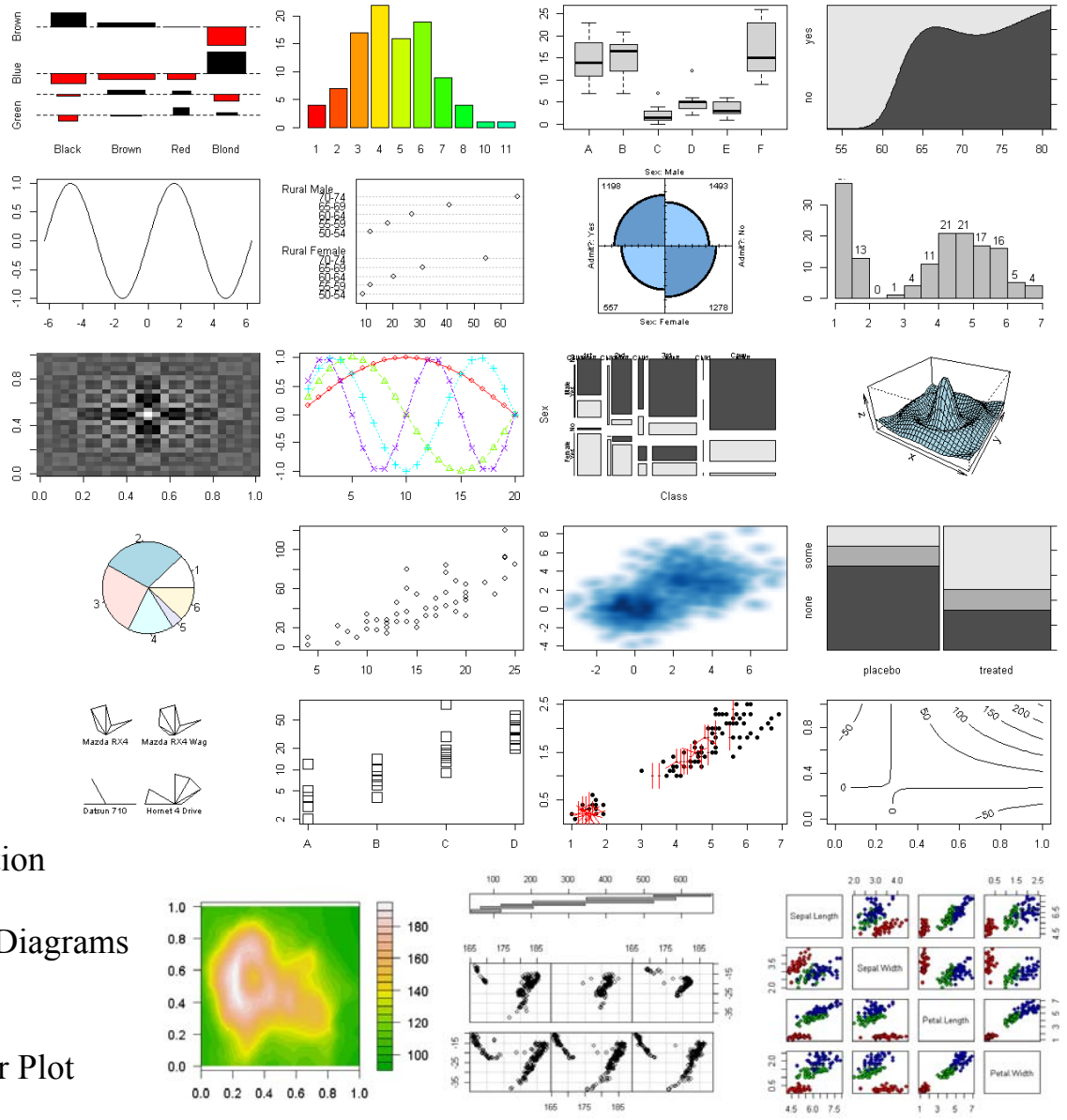
Graphical Perception Tasks.
Ordered from the most accurate to the least accurate (Jacoby, 1997)



graphics: The R Graphics Package

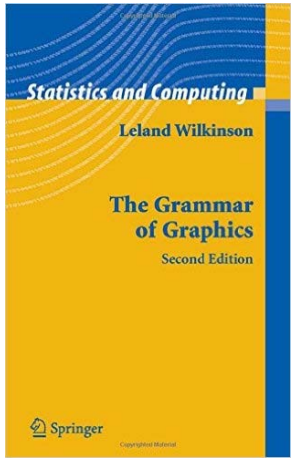
Plots:

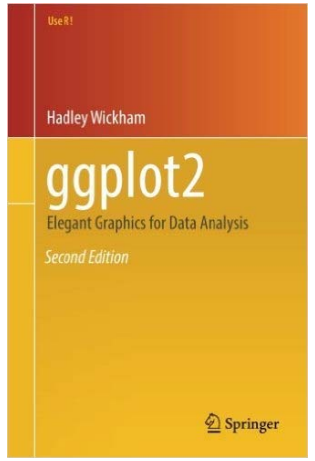
assocplot: Association Plots
barplot: Bar Plots
boxplot: Box Plots
cdplot: Conditional Density Plots
contour: Display Contours
coplot: Conditioning Plots
curve: Draw Function Plots
dotchart: Cleveland's Dot Plots
filled.contour: Level (Contour) Plots
fourfoldplot: Fourfold Plots
hist: Histograms
image: Display a Color Image
matplot: Plot Columns of Matrices
mosaicplot: Mosaic Plots
pairs: Scatterplot Matrices
persp: Perspective Plots
pie: Pie Charts
plot: Generic X-Y Plotting
smoothScatter: Scatterplots with Smoothed Densities Color Representation
spineplot: Spine Plots and Spinograms
stars: Star (Spider/Radar) Plots and Segment Diagrams
stem: Stem-and-Leaf Plots
stripchart: 1-D Scatter Plots
sunflowerplot: Produce a Sunflower Scatter Plot



ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics

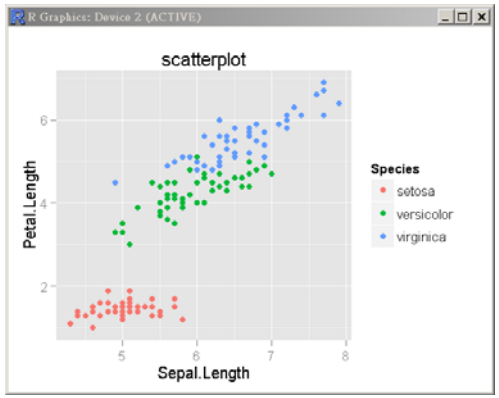
- Hadley Wickham, ggplot2: Elegant Graphics for Data Analysis: http://ggplot2.org/

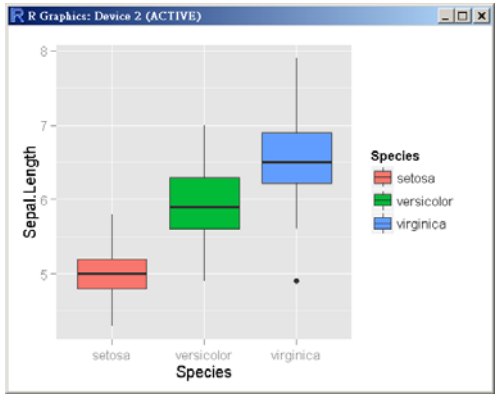


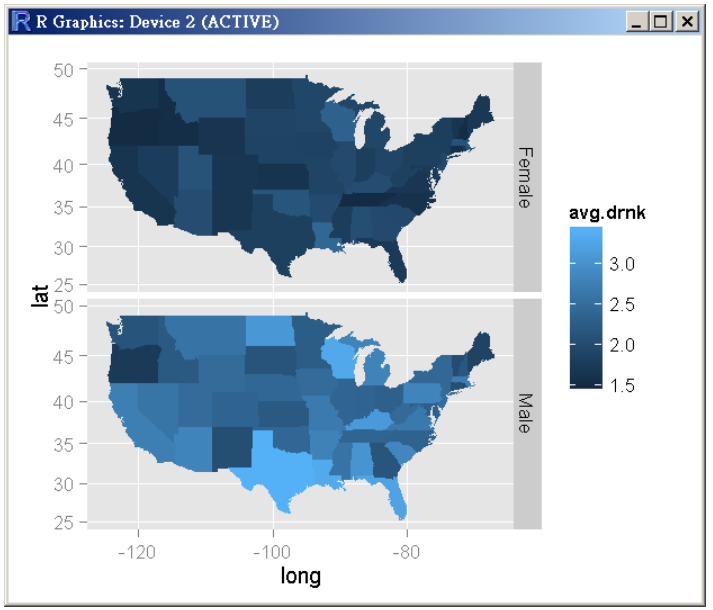



```
qplot(x, y = NULL, ..., data, facets = NULL, margins = FALSE,
  geom = "auto", stat = list(NULL), position = list(NULL),
  xlim = c(NA,NA), ylim = c(NA, NA), log = "", main = NULL,
  xlab = deparse(substitute(x)),
  ylab = deparse(substitute(y)), asp = NA)
```

```
library(ggplot2)
qplot(Sepal.Length, Petal.Length, geom="point",
 data=iris, colour = Species, main="scatterplot")
qplot(Species, Sepal.Length, geom="boxplot",
 fill=Species, data=iris)
```







ggplot2 Version of Figures in Lattice:
https://learnr.files.wordpress.com/2009/08/latbook.pdf

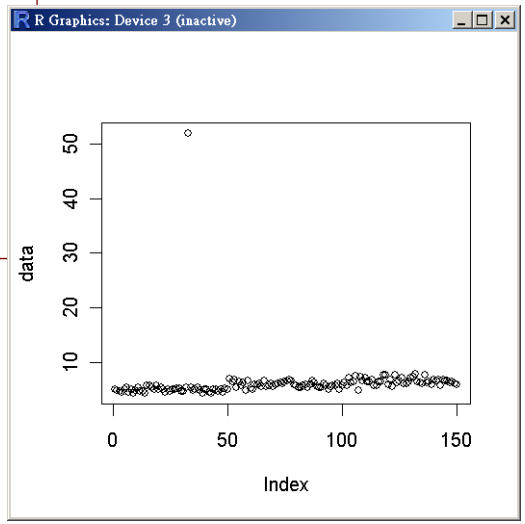
http://www.youtube.com/watch?v=HeqHMM4ziXA
http://www.youtube.com/watch?v=n8kYa9vu1l8

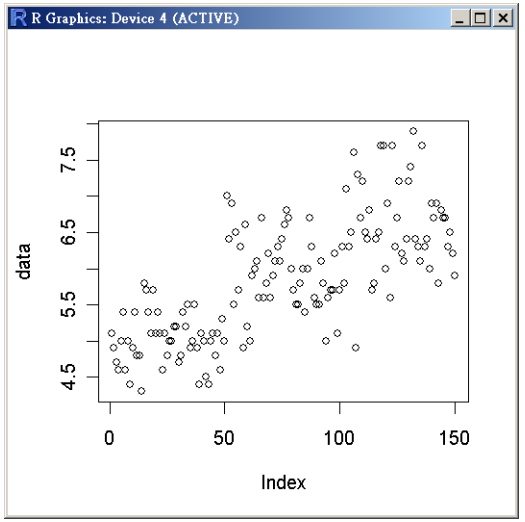
索引圖 (Index Plot)

- Index plot takes a single argument which is a continuous variable and plots the values on the y axis, with the x coordinate determined by the position of the number in the vector.

- Useful for error checking.

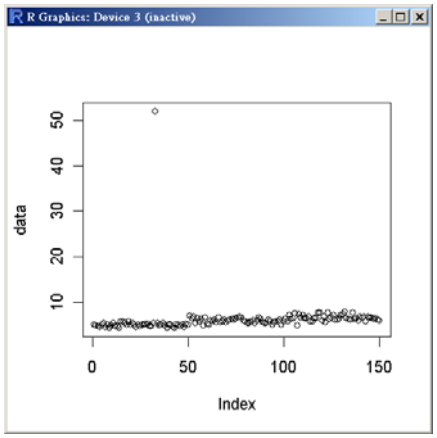
```
> data <- iris[,1]
> data[33] <- data[33]*10
> plot(data)
> ind <- which(data>15)
> data[ind]
> data[ind] <- 5.2
> windows()
> plot(data)
```

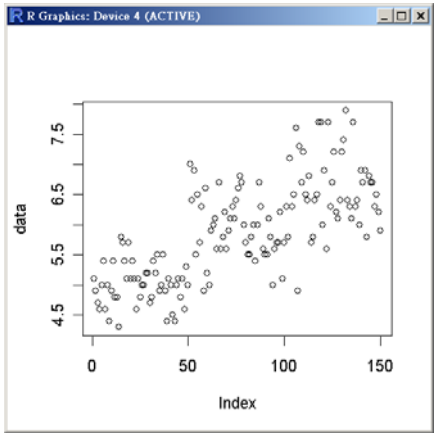


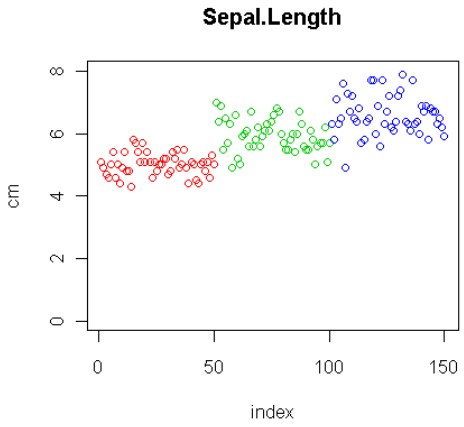


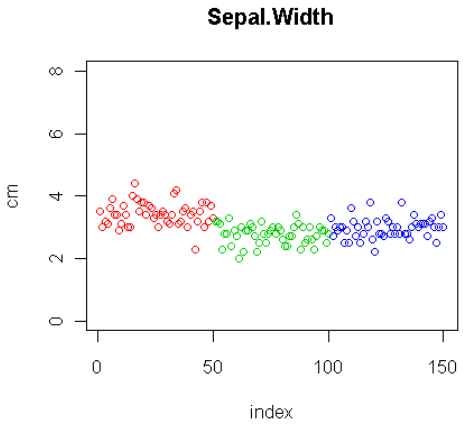
Index Plot

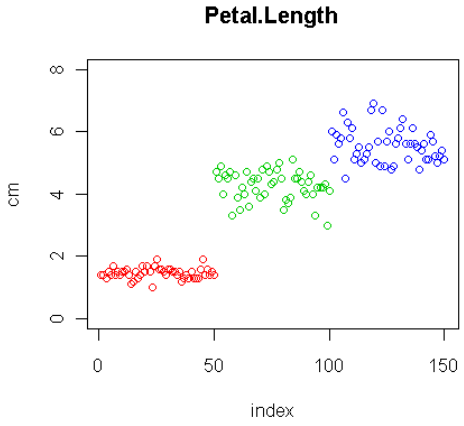
- Index plot takes a single argument which is a continuous variable and plots the values on the y axis, with the x coordinate determined by the position of the number in the vector.
- Useful for error checking.

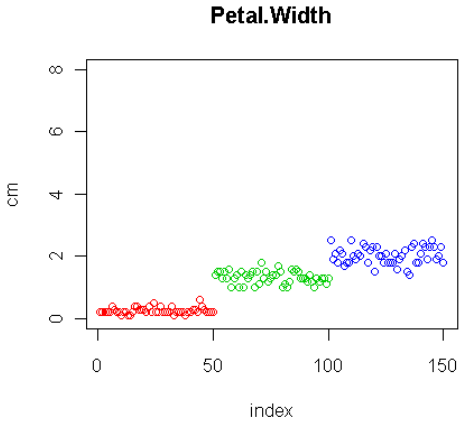








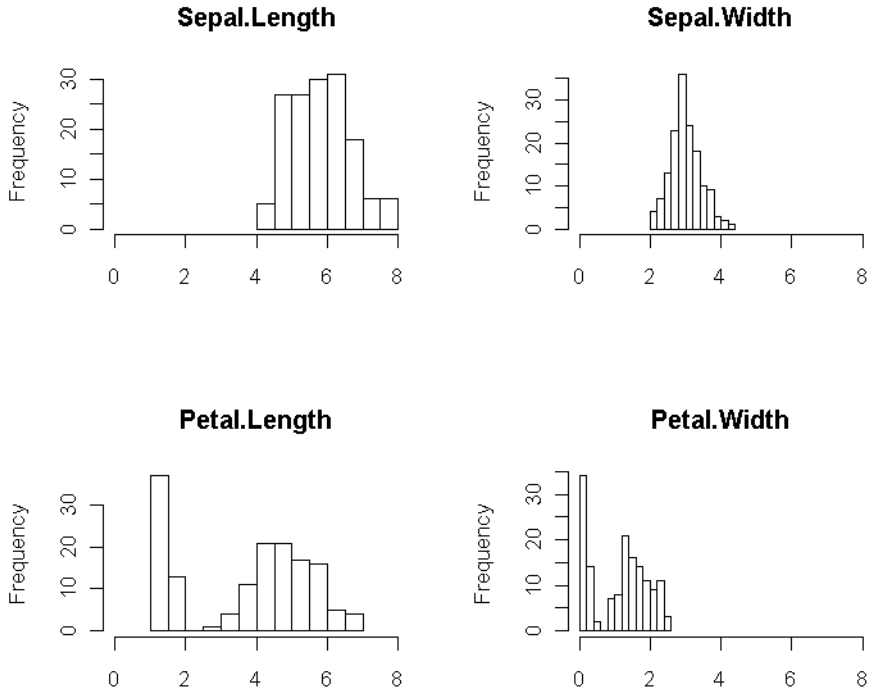


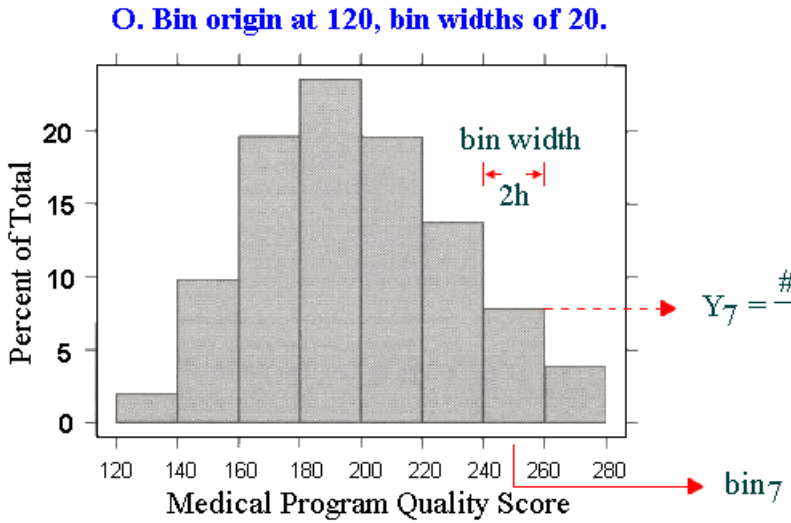


直方圖 (Histogram) (1/3)

The histogram shows:

1. center of the data (location)
2. spread of the data (scale)
3. skewness of the data
4. presence of outliers
5. presence of multiple modes in the data.





$$Y_7 = \frac{\#\text{ observations within bin}_7}{(2h)\,n}$$

Changes in bin origin and bin widths affect the shape of the histogram

直方圖 (Histogram) (2/3)

- 1/2h adjusts the height of each bar so that the total area enclosed by the entire histogram is 1.
- The area covered by each bar can be interpreted as the probability of an observation falling within that bar.

Disadvantage for displaying a variable's distribution:

- selection of origin of the bins.
- selection of bin widths.
- the very use of the bins is a distortion of information because any data variability within the bins cannot be displayed in the histogram.

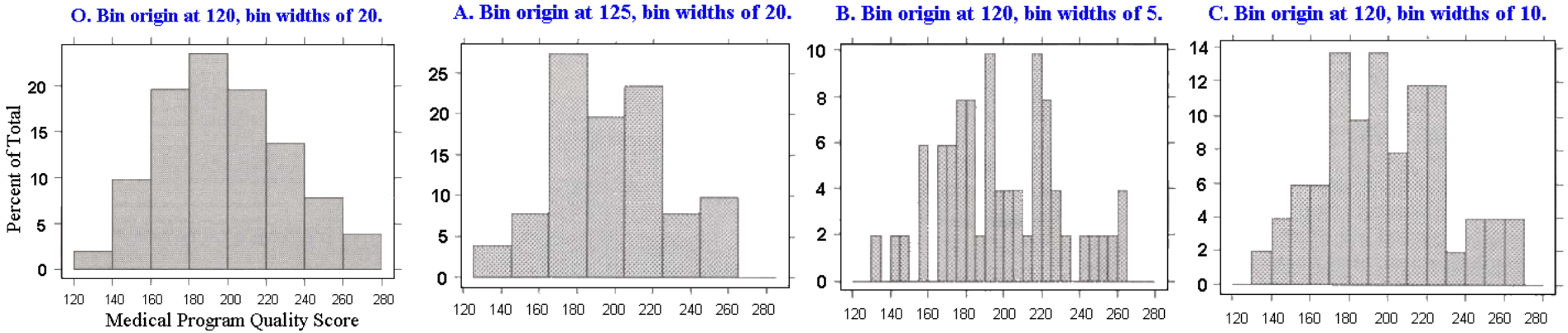
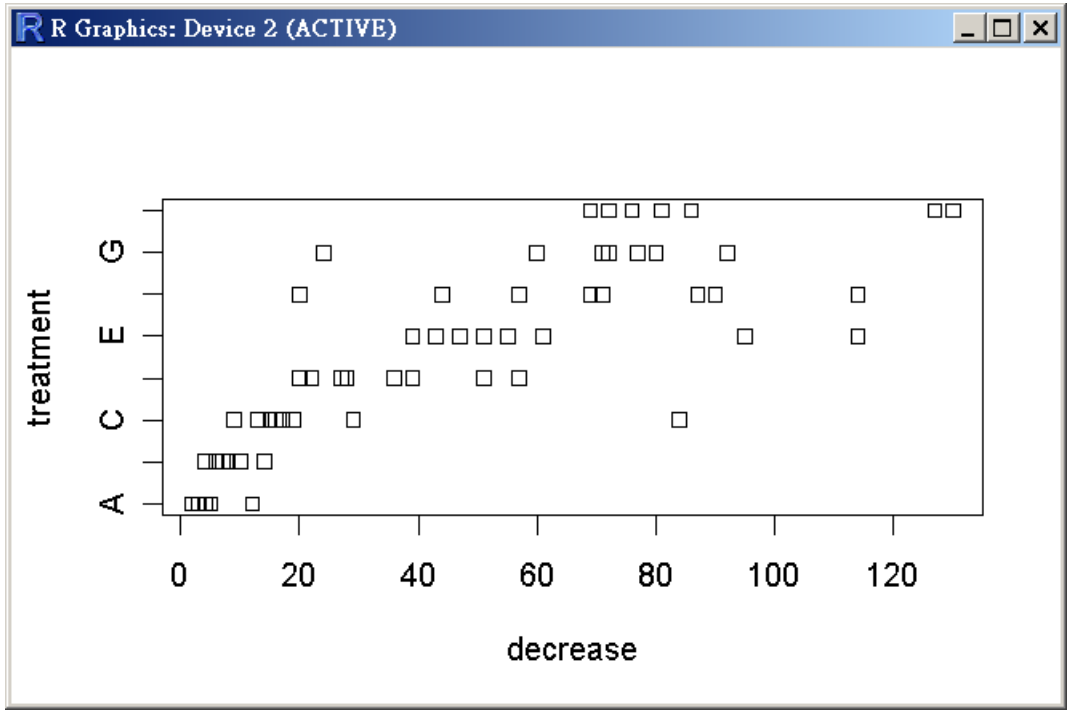


Figure Sources: Jacoby (1997).

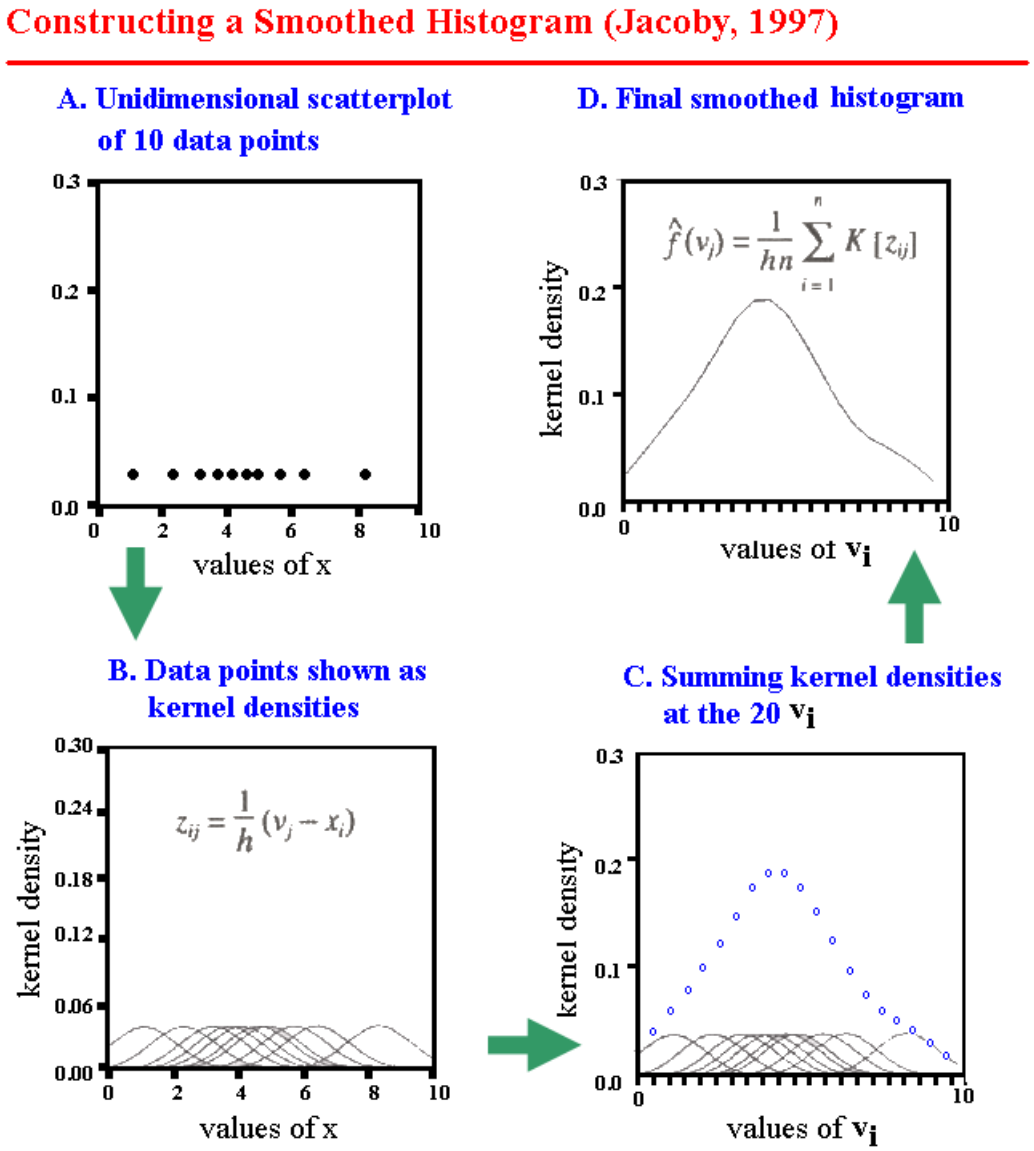
帶形圖 (Stripchart)

```
> attach(OrchardSprays)
> names(OrchardSprays)
[1] "decrease"  "rowpos"    "colpos"    "treatment"
> OrchardSprays[1:5,]
> stripchart(decrease~treatment, xlab="decrease", ylab="treatment")
```



Density Plots (Smoothed Histograms) (1/3)

- Smoothed histograms overcome some of the disadvantages caused by the arbitrary, discrete bins used in traditional histogram.
- The relative height of the smooth curve corresponds to the local density.
- The overall height is adjusted so that the total area under the curve is approximately equal to 1.
- The area under the curve between any two points along the horizontal scale can be interpreted as the probability that an observation falls within that interval of data values.

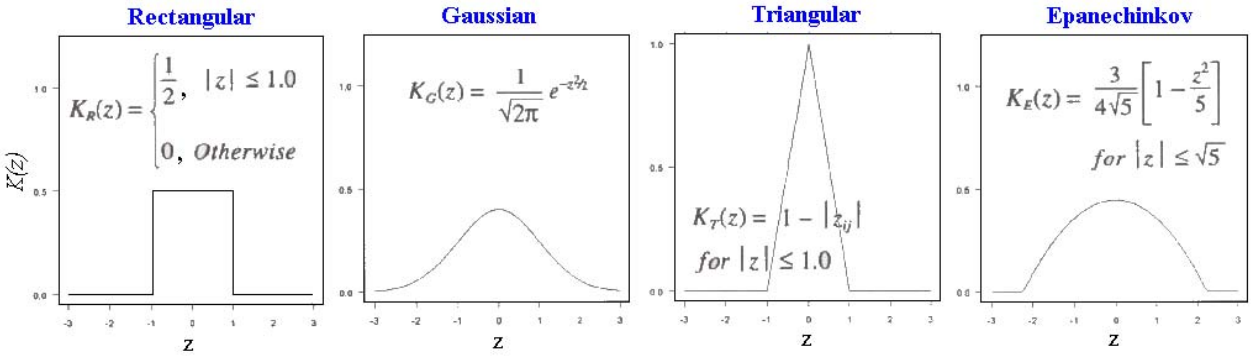


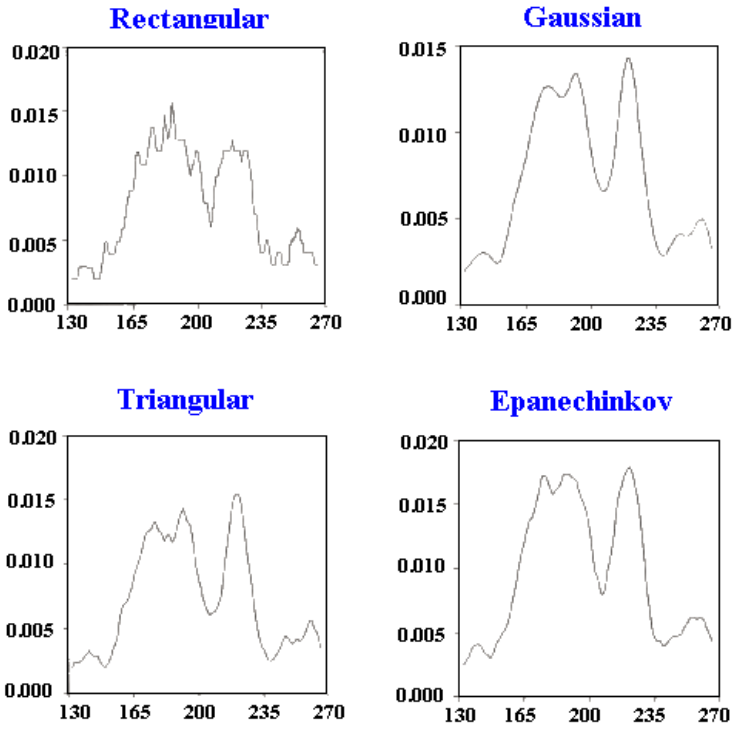
Density Plots (2/3)

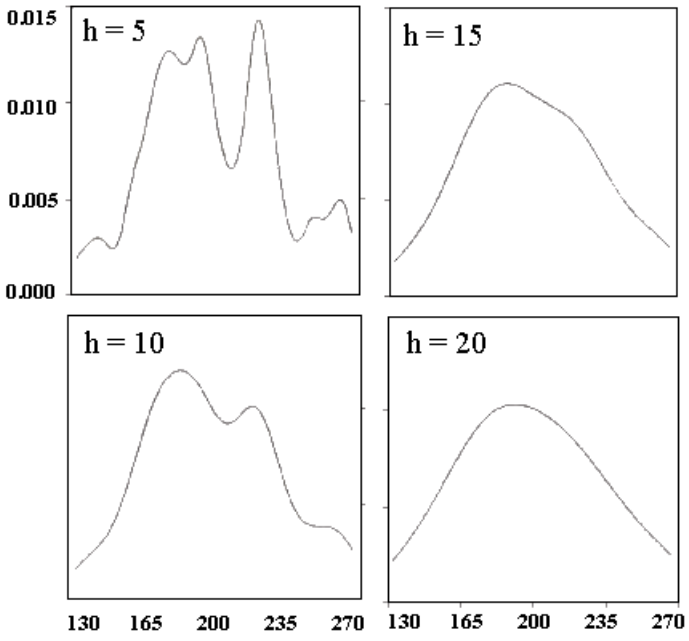
- Selection of kernels
- Selection of bandwidth

$$\hat{f}(v_j) = \frac{1}{hn}\sum_{i=1}^{n} K\left[z_{ij}\right]$$

$$z_{ij} = \frac{1}{h}(v_j - x_i)$$







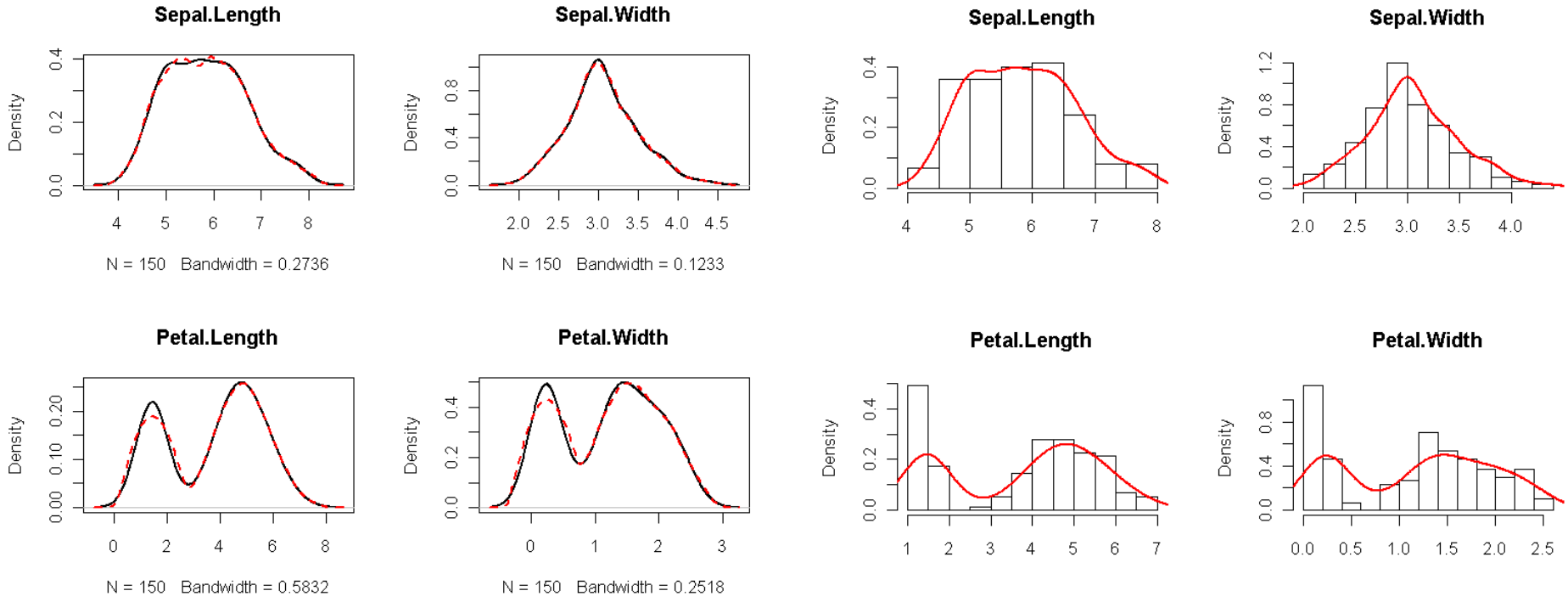
Figures modified from Jacoby (1997)

Density Plots (3/3)

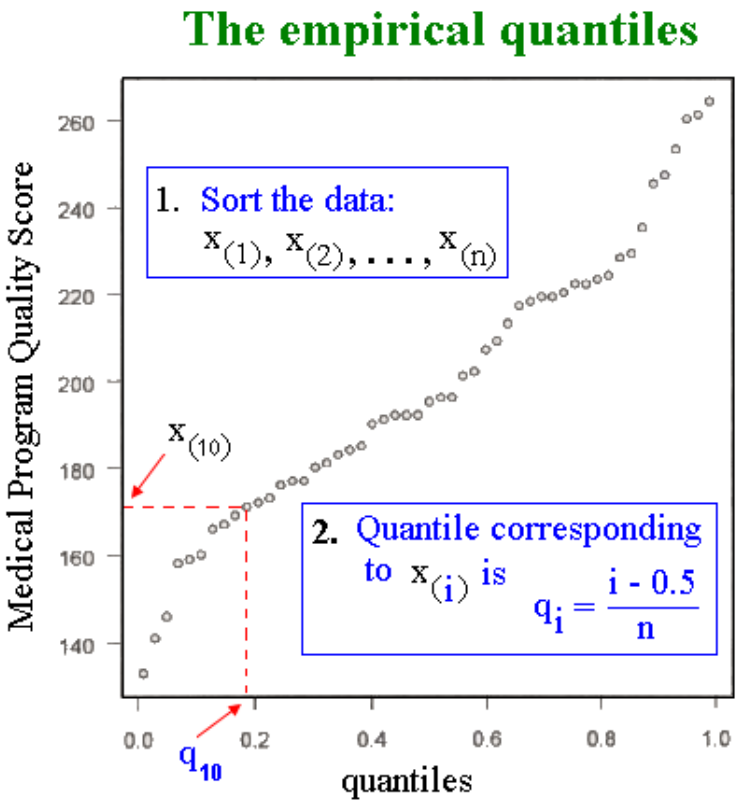
```
density(x, bw = "nrd0", adjust = 1,
        kernel = c("gaussian", "epanechnikov", "rectangular",
                   "triangular", "biweight",
                   "cosine", "optcosine"),
        weights = NULL, window = kernel, width,
        give.Rkern = FALSE,
        n = 512, from, to, cut = 3, na.rm = FALSE, ...)
```

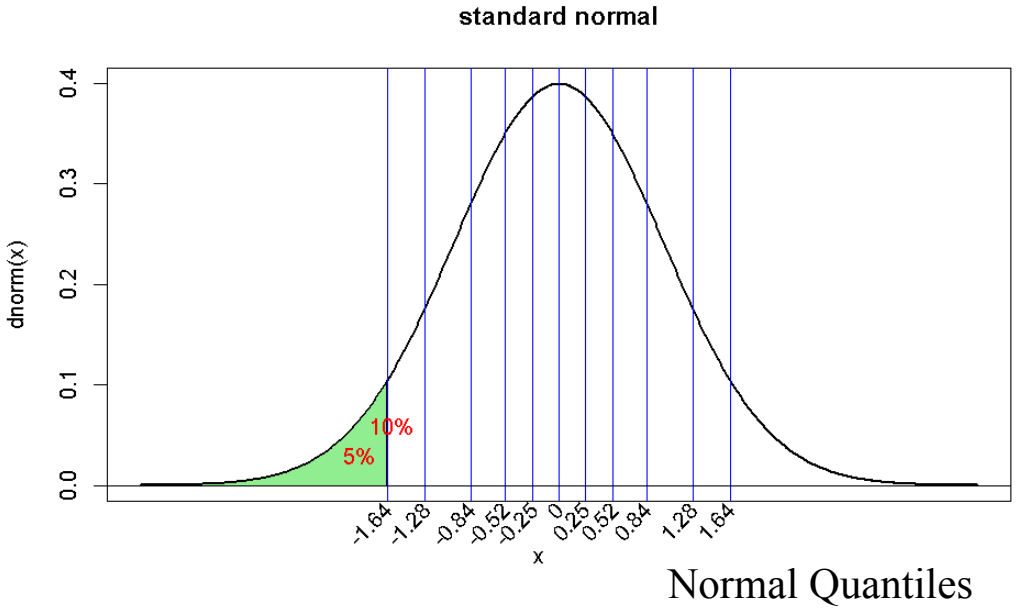
gaussian
epanechnikov

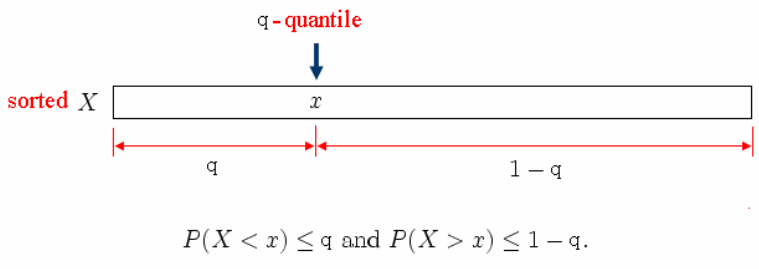
```
> plot(density(iris$Sepal.Length))
```



Quantile Plots







- 0.5 is subtracted from each i value to avoid extreme quantiles of exactly 0 or 1.
- The latter would cause problems if empirical quantiles were to be compared against quantiles derived from a theoretical asymptotic distribution such as the normal.
- This adjustment has no effect on the shape of any graphical display.

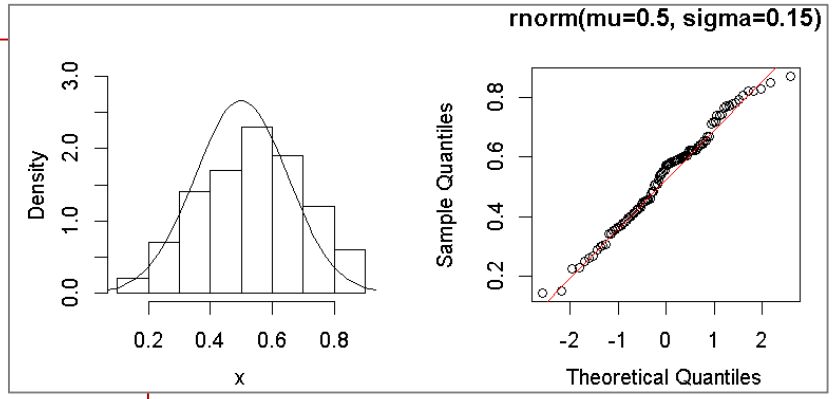
Quantile-Quantile Plots

- The quantile-quantile (Q-Q) plot is used to determine if two data sets come from populations with a common density.
- Q-Q plots are sometimes called probability plots, especially when data are examined against a theoretical density.

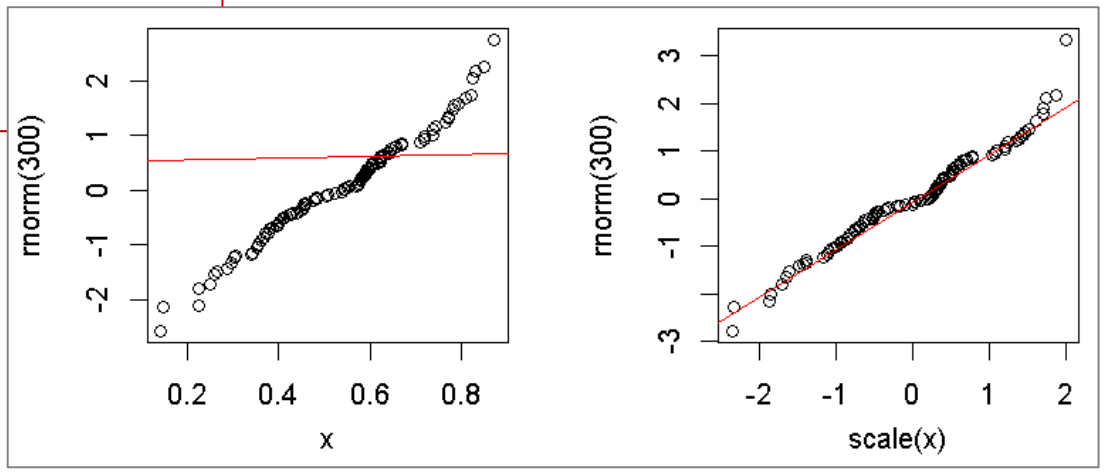
- **`qqnorm()`**: produces a normal QQ plot of the values in sample
- **`qqline()`**: adds a line which passes through the first and third quartiles.
  - Use the diagonal line would not make sense because the first axis is scaled in terms of the theoretical quantiles of a N (0,1) distribution.
  - Using the first and third quartiles to set the line gives a robust approach for estimating the parameters of the normal distribution, when compared with using the empirical mean and variance, say.
  - Departures from the line (except in the tails) are indicative of a lack of normality.
- **`qqplot()`**: qqplot produces a QQ plot of two datasets.

qqnorm, qqline, qqplot

```
> par(mfrow = c(1, 2))
> set.seed(12345);
> n <- 100; mu <- 0.5; sigma <- 0.15
> x <- rnorm(n, mu, sigma)
> hist(x, freq=FALSE, ylim=c(0, 3), main="")
> y <- seq(0, 1, length = n)
> lines(y, dnorm(y, mu, sigma), type = 'l')
> qqnorm(x, main = "rnorm(mu=0.5, sigma=0.15)");
> qqline(x)
```




```
> qqplot(x, rnorm(300))
> qqline(x, col = 2)
> qqplot(scale(x), rnorm(300))
> qqline(scale(x), col = 2)
```

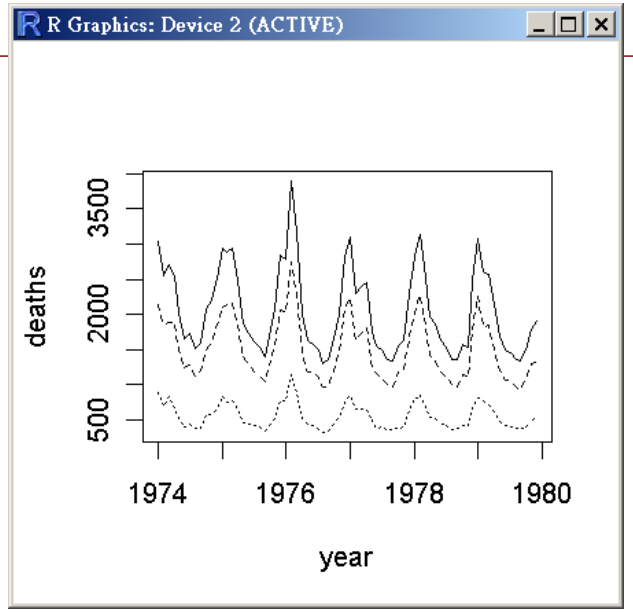


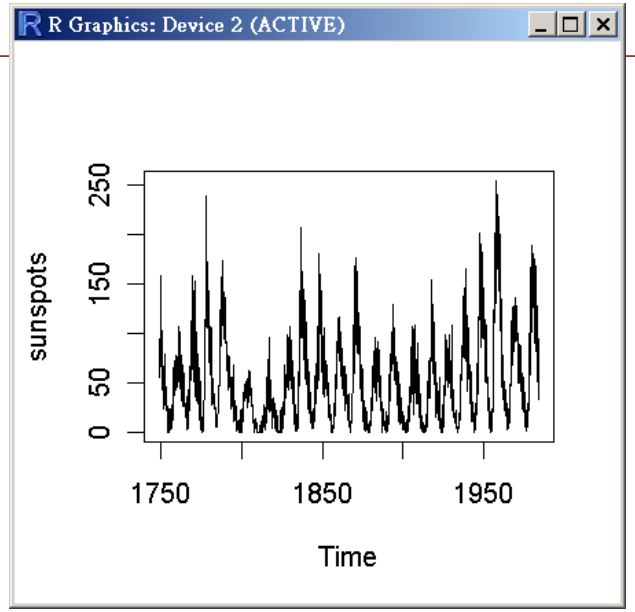
時間序列 (Time Series Plots)

- **ts.plot**: simple vector of numbers
- **plot.ts** works for plotting objects inheriting from class ts.

```
> data(UKLungDeaths) # total, male, female death
> ts.plot(ldeaths, mdeaths, fdeaths, xlab="year", ylab="deaths", lty=c(1:3))

> data(sunspots)
> plot(sunspots) # sunspots is ts class
> class(sunspots)
[1] "ts"
> is.ts(sunspots)
[1] TRUE
```

```
data(UKgas)
attach(UKgas)
names(UKgas)
```

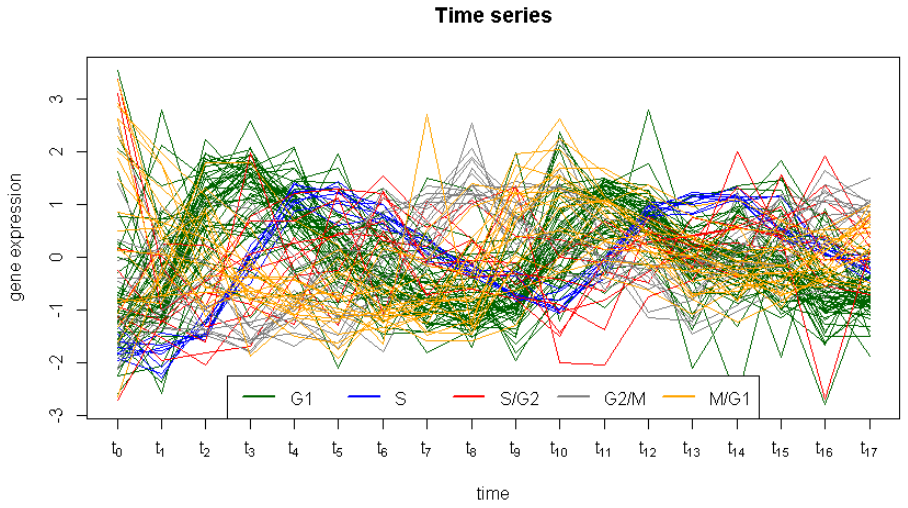




Time Series Plots

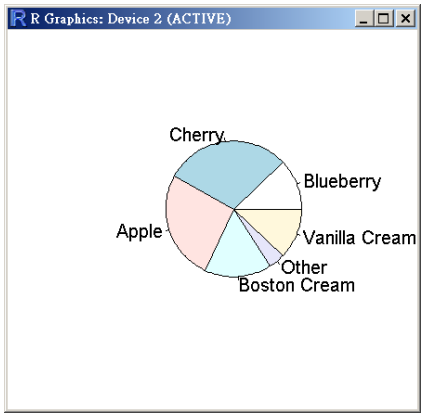
```
cell.raw <- read.table("trad_alpha103.txt", row.names=1, header=T)
head(cell.raw
cell.xdata <- t(scale(t(cell.raw[,2:19]), center=T, scale=T))
y.C <-  as.integer(cell.raw[,1])
table(y.C)
no.cluster <- length(unique(y.C))
p <- ncol(cell.raw) - 1
cellcycle.color <- c("darkgreen", "blue", "red", "gray50", "orange")
ycolors <- cellcycle.color[y.C+1]
my.pch <- c(1:no.cluster)[y.C+1]
phase <- c("G1", "S", "S/G2", "G2/M", "M/G1")
matplot(t(cell.xdata), lty=1, type = "l", ylab="gene expression",
            col=ycolors, xlab="time", main="Time series", xaxt="n")
time.label <- parse(text=paste("t[",0:p,"]",sep=""))
axis(1, 1:(p+1), time.label)
legend("bottom", legend=phase, col=cellcycle.color, lty=1, horiz = T, lwd=2)
```

	A	B	C	D	E	F	G
1	UID	Phase	alpha0	alpha7	alpha14	alpha21	alpha28
2	YAR007C_G1	0	-0.48	-0.42	0.87	0.92	0.67
3	YBL035C_G1	0	-0.39	-0.58	1.08	1.21	0.52
4	YBR023C_G1	0	0.87	0.25	-0.17	0.18	-0.13
5	YBR067C_G1	0	1.57	1.03	1.22	0.31	0.16
6	YBR088C_G1	0	-1.15	-0.86	1.21	1.62	1.12
7	YBR278W_G1	0	0.04	-0.12	0.31	0.16	0.17
8	YCL055W_G1	0	2.95	0.45	-0.40	-0.66	-0.59
9	YDL003W_G1	0	-1.22	-0.74	1.34	1.50	0.63
10	YDL055C_G1	0	-0.73	-1.06	-0.79	-0.02	0.16
11	YDL102W_G1	0	-0.58	-0.40	0.13	0.58	-0.09
12	YDL164C_G1	0	-0.50	-0.42	0.66	1.05	0.68
13	YDL197C_G1	0	-0.86	-0.29	0.42	0.46	0.30
14	YDL227C_G1	0	-0.16	0.29	0.17	-0.28	-0.02
15	YDR052C_G1	0	-0.36	-0.03	-0.03	-0.08	-0.23

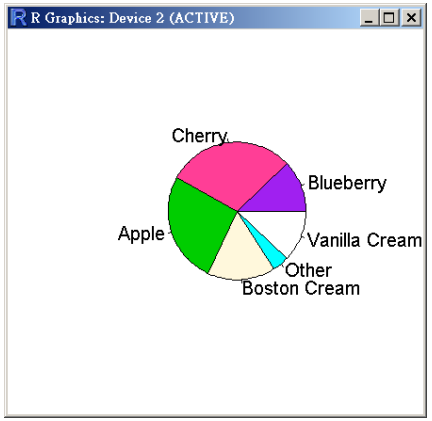


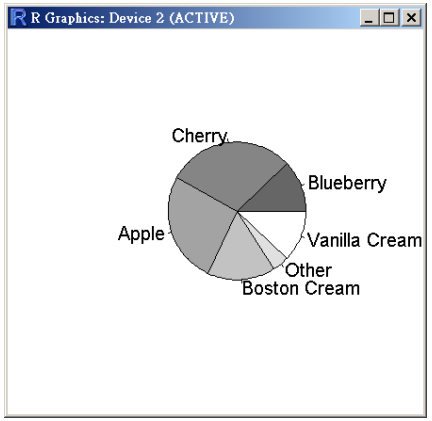
圓餅圖 (Pie Charts)

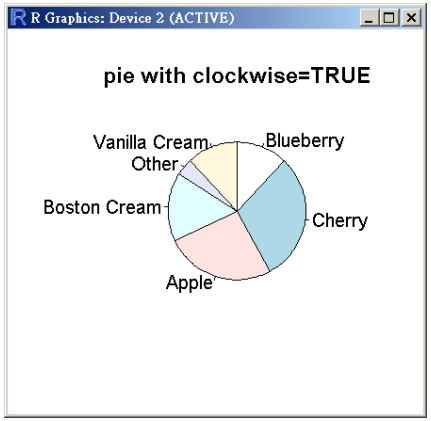
```
> pie.sales <- c(0.12, 0.3, 0.26, 0.16, 0.04, 0.12)
> sum(pie.sales)
> names(pie.sales) <- c("Blueberry", "Cherry",
    "Apple", "Boston Cream", "Other", "Vanilla Cream")

> pie(pie.sales) # default colours
```

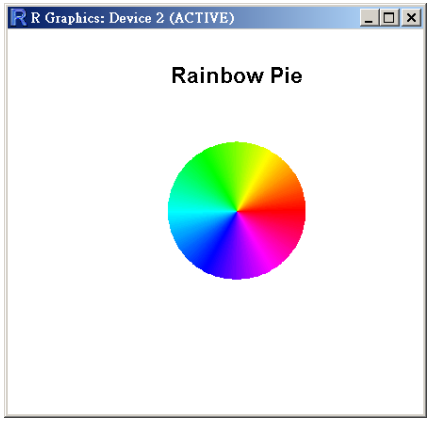



```
> pie(pie.sales, col = c("purple", "violetred1", "green3", "cornsilk", "cyan", "white"))
> pie(pie.sales, col = gray(seq(0.4,1.0,length=6)))
> pie(pie.sales, clockwise=TRUE, main="pie with clockwise=TRUE")
> pie(rep(1,200), labels="", col=rainbow(200), border=NA, main = "Rainbow Pie")
```



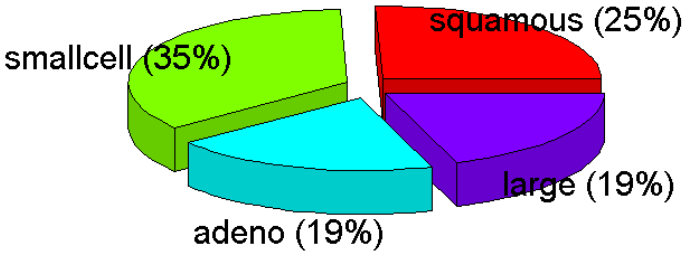


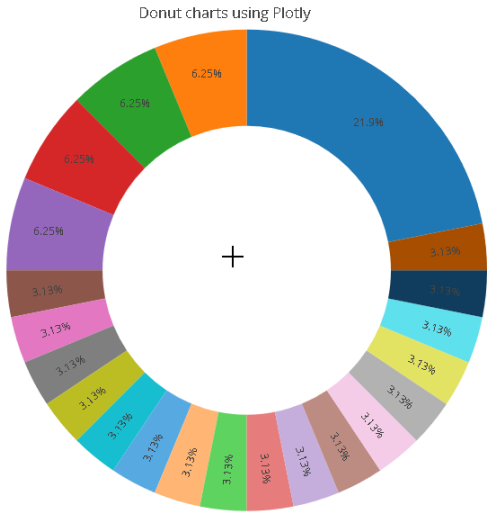




Pie Charts and Their Variants

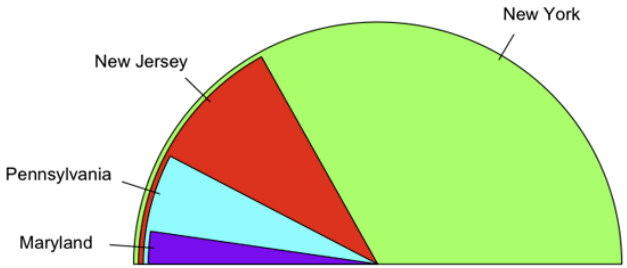
```
> library(plotrix)
> library(survival)
> head(veteran)
  trt celltype time status karno diagtime age prior
1   1 squamous   72      1    60         7  69     0
2   1 squamous  411      1    70         5  64    10
3   1 squamous  228      1    60         3  38     0
4   1 squamous  126      1    60         9  63    10
5   1 squamous  118      1    70        11  65    10
6   1 squamous   10      1    20         5  49     0
> slices <- summary(veteran$celltype)
> p <- floor(100*slices/sum(slices))
> pie3D(slices, labels=paste0(names(slices), " (",p, "%)"), explode=0.1)
```



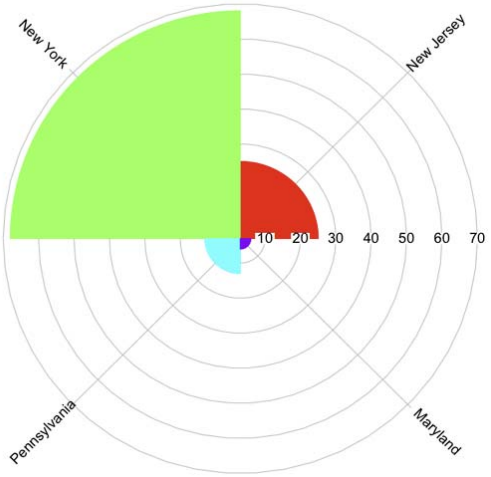


https://plot.ly/r/pie-charts/

fan.plot



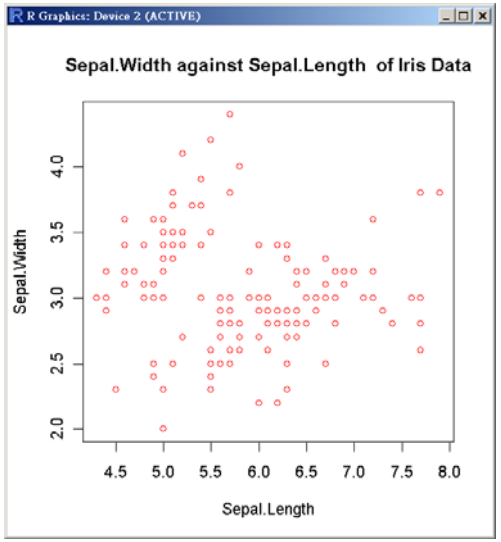
radial.pie



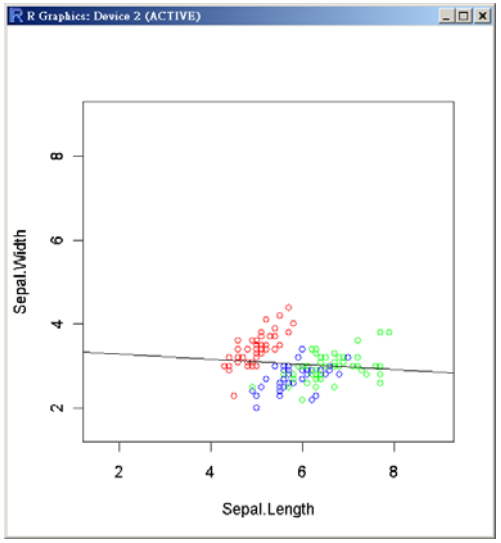
散佈圖 (Scatterplot)

`plot(x, y)` 或 `plot(y~x)`

```
> xlab <- names(iris)[1]
> ylab <- names(iris)[2]
> title <- paste(ylab, "against", xlab, " of Iris
    Data")
> x <- iris[,1]
> y <- iris[,2]
> plot(x, y, col="red", xlab=xlab, ylab=ylab,
    main=title)
```



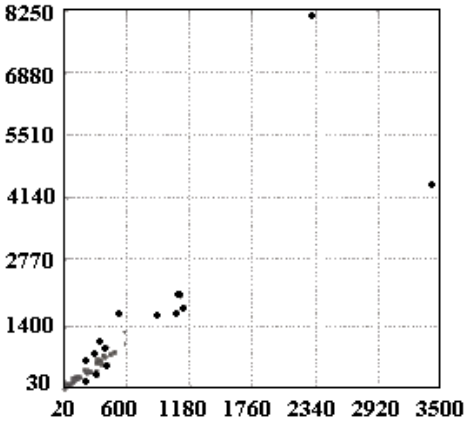

```
> range(x)
> range(y)
> plot(y~x, xlab=xlab, ylab=ylab, xlim=c(1.5,9),
     ylim=c(1.5,9), type="n")
> points(x[1:50], y[1:50], col="red")
> points(x[51:100], y[51:100], col="blue")
> points(x[101:150], y[101:150], col="green")
> abline(lm(y~x))
```



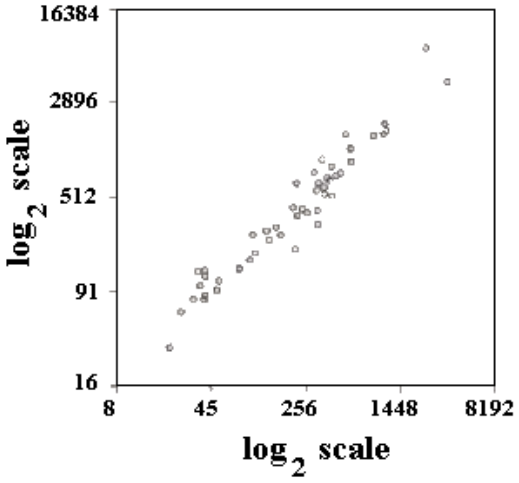
Extensions of Scatterplots

With a smoothing curve

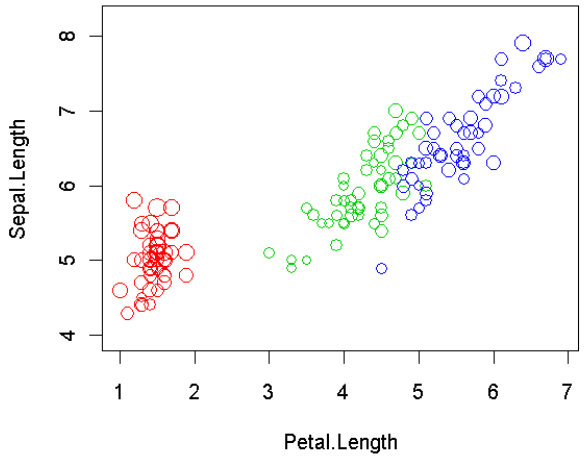
Poorly-Constructed



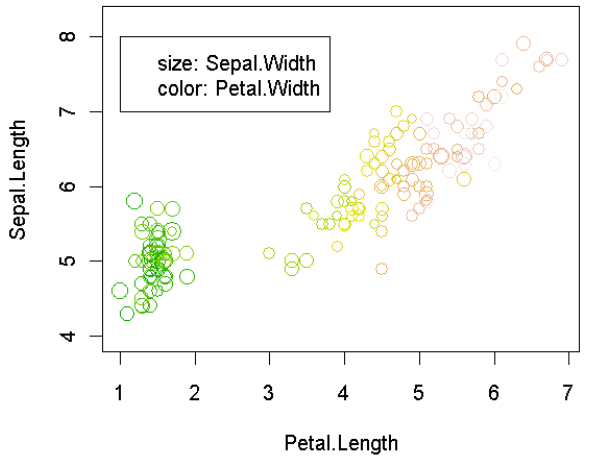
Better



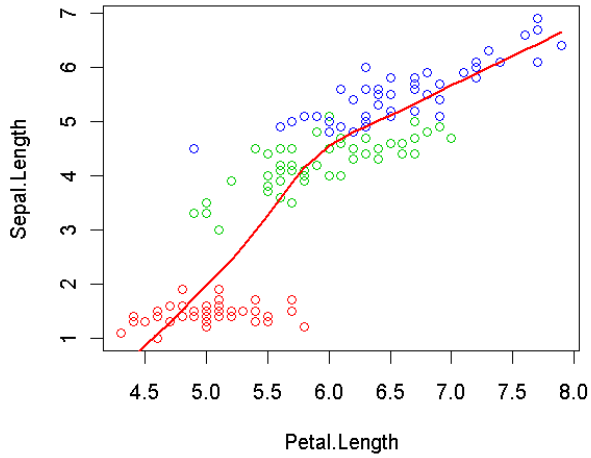
Bubbleplot: Sepal.Width



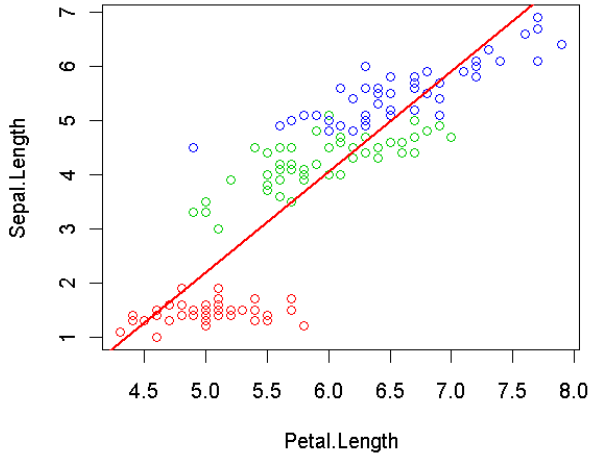
Color Bubbleplot



LOWESS



Simple Linear Regression



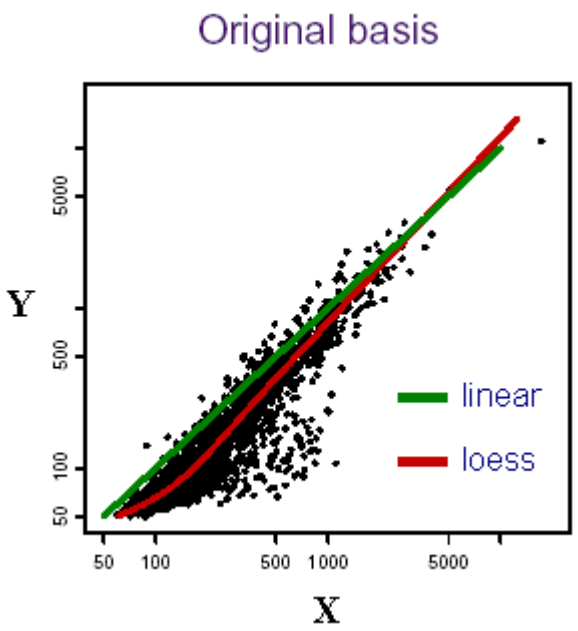
MA plot
Scatterplot for Gene Expression Data

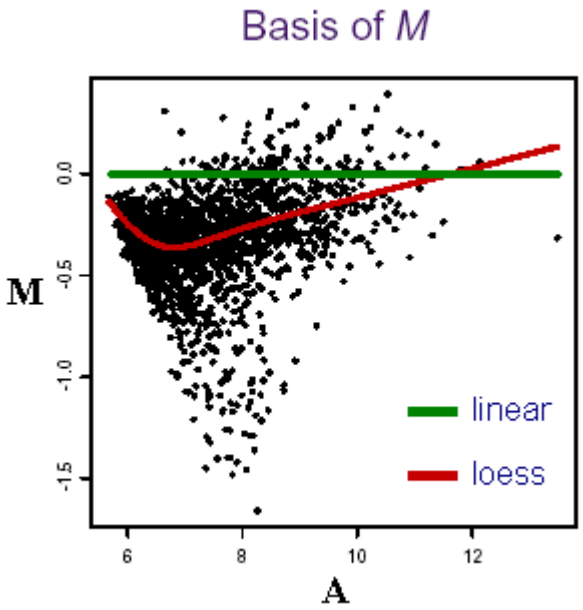
- **MA plots** can show the intensity-dependent ratio of raw microarray data.

Oligo	cDNA
$X = PM_1$, $Y = PM_2$	X= Cy3, Y= Cy5
$X = PM_1 - MM_1$, $Y = PM_2 - MM_2$	

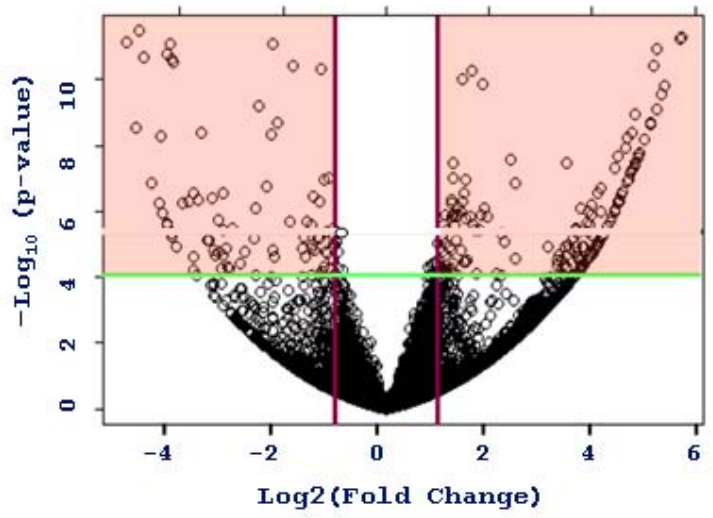
$$M = \log_2\left(\frac{Y}{X}\right)$$

$$A = \frac{1}{2}\log_2(XY)$$





Volcano Plot



- A volcano plot is a heuristic device that arranges genes along dimensions of biological and statistical significance.
- A volcano plot is helpful in identifying significance and magnitude of change in expression of a set of genes between two conditions.

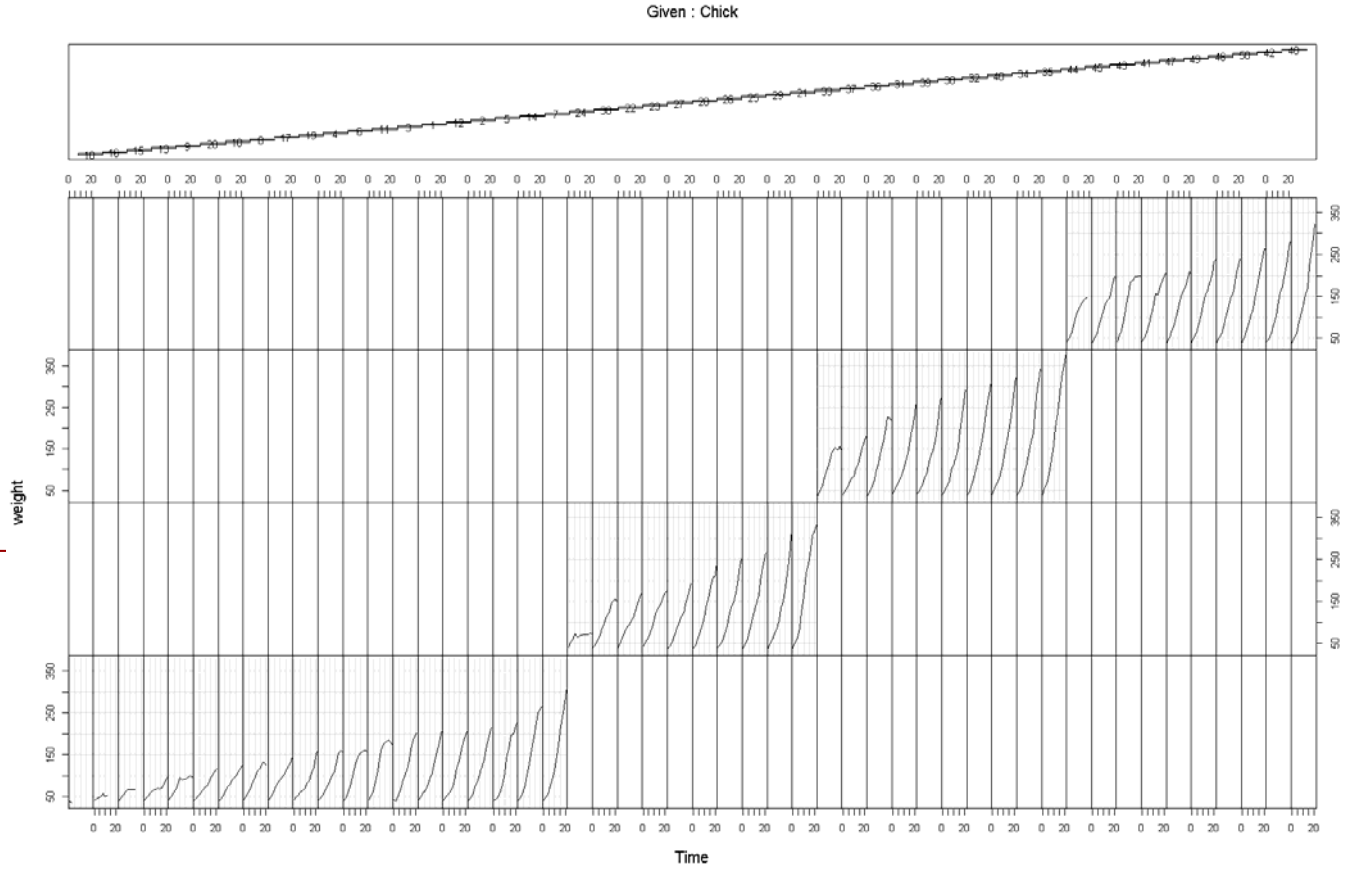
- A volcano plot displays the negative log of p-values from a t-test on one axis and the log2 of change between two conditions on the other axis on the scatterplot view.

- The researcher can then make judgments about the most promising candidates for follow-up studies, by trading off both these criteria by eye.

coplot {graphics}: Conditioning Plots

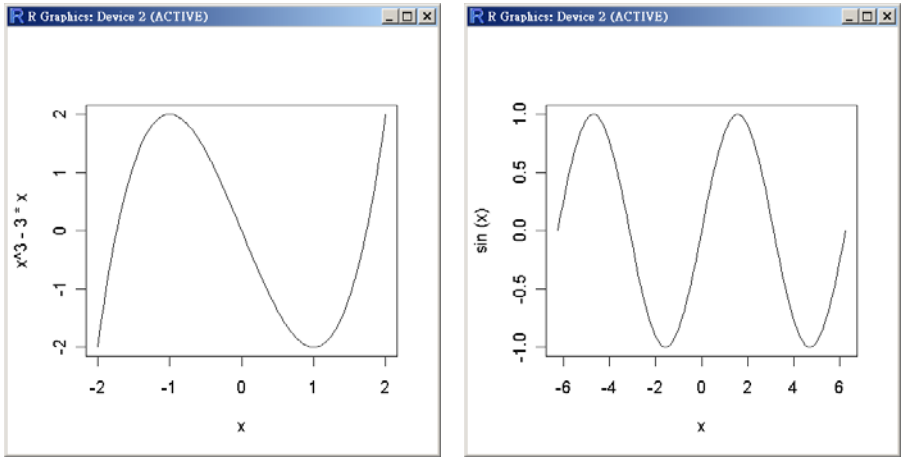
```
> head(ChickWeight, 3)
  weight Time Chick Diet
1     42    0     1    1
2     51    2     1    1
3     59    4     1    1
> coplot(weight ~ Time | Chick * Diet, type = "l", data = ChickWeight)
```

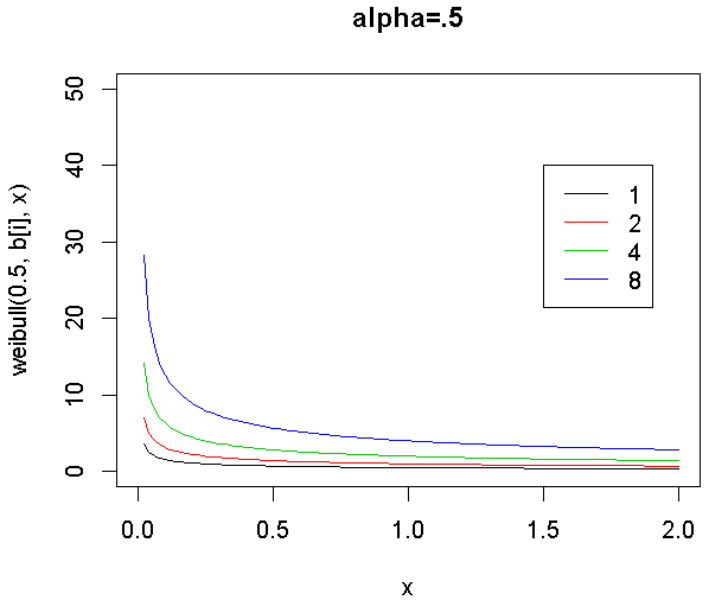
```
> table(ChickWeight$Chick, ChickWeight$Diet)

      1  2  3  4
  18  2  0  0  0
  16  7  0  0  0
  15  8  0  0  0
  13 12  0  0  0
  9  12  0  0  0
  20 12  0  0  0
  10 12  0  0  0
  8  11  0  0  0
  17 12  0  0  0
  19 12  0  0  0
  4  12  0  0  0
...
```



curve {graphics}: 畫2D數學函數圖

```
> curve(x^3-3*x, -2, 2)
> curve(sin, -2*pi, 2*pi)
```





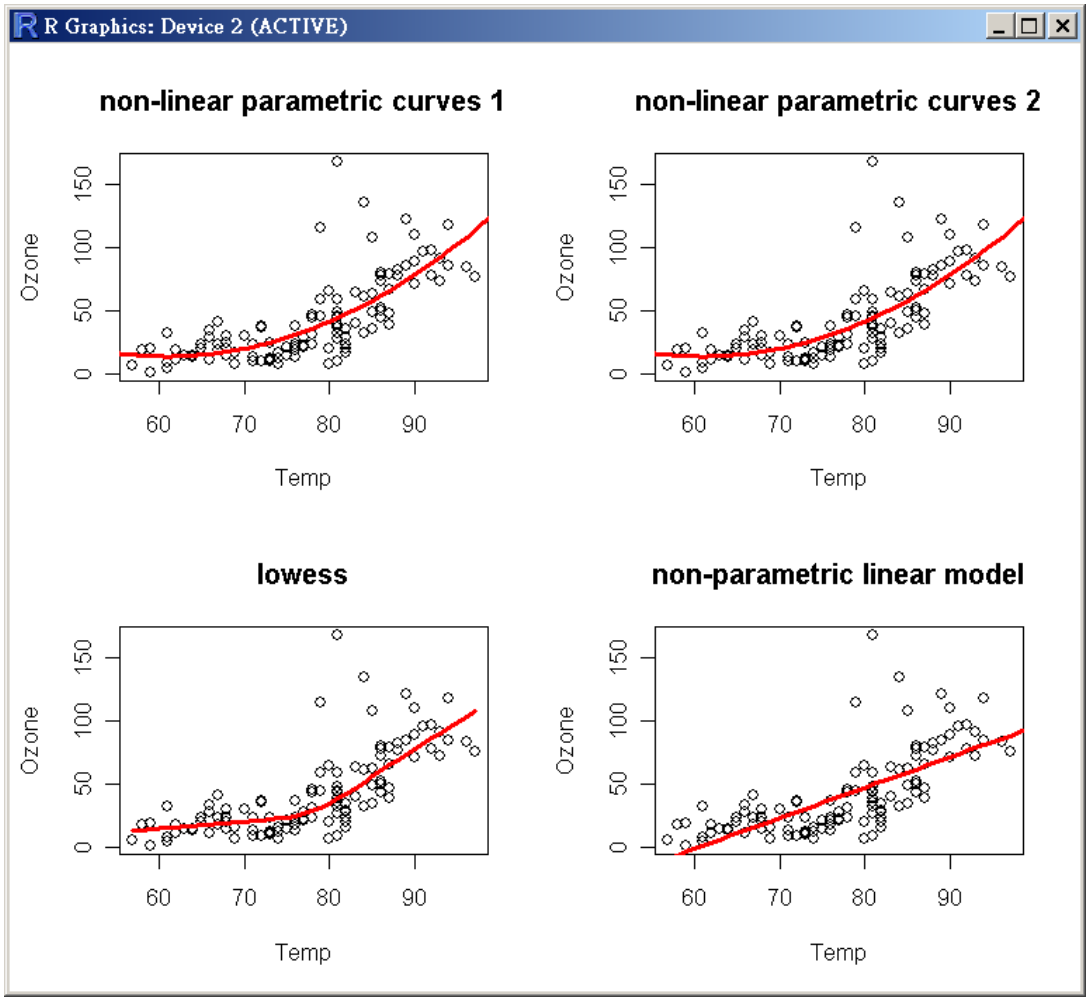

```
> x <- seq(-2, 2, 0.01)
> y <- x^3-3*x
> plot(x, y, type="l")
```

```
> weibull <- function(alpha, beta, x){
+   alpha * beta * (x^(alpha-1))
+ }
>
> b <- c(1, 2, 4, 8)
> for(i in 1:length(b)) {
+   curve(weibull(0.5, b[i], x), from=0, to=2,
+         add=(i!=1),
+         col=i, ylim=c(0, 50), main="alpha=.5")
+ }
>
> legend(1.5, 40, legend=b, col=1:length(b), lty=1)
```

曲線配適 (Fitting Curves)

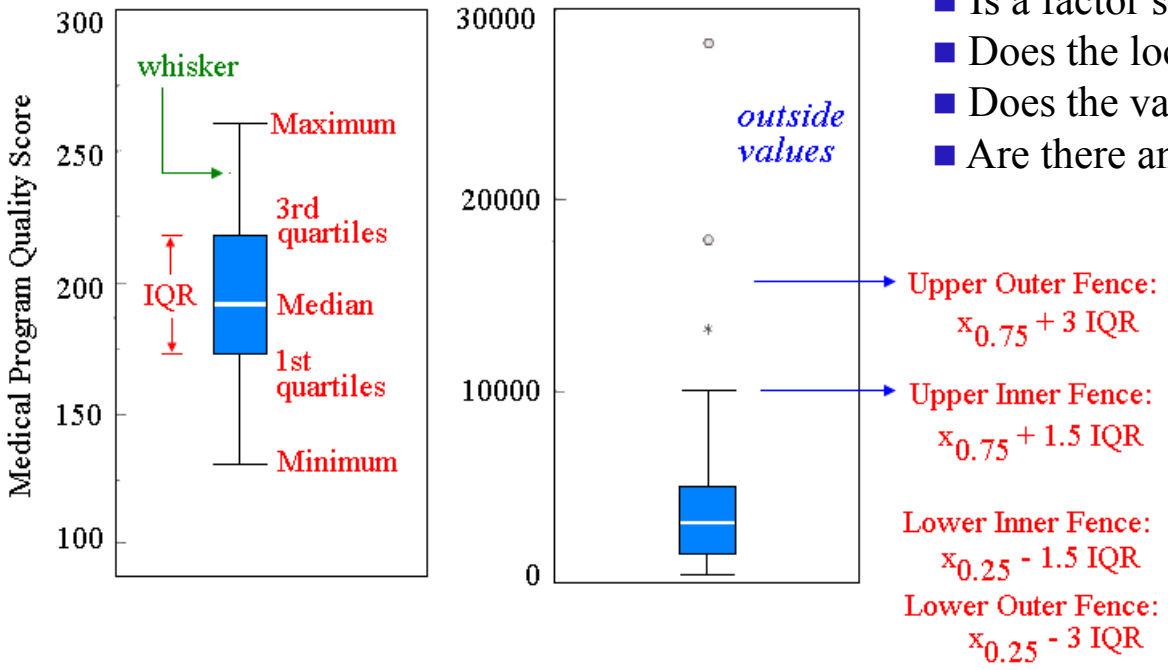
Example Methods

- non-linear parametric curves
- lowess (a non-parametric curve fitter)
- loess (a modelling tool)
- gam (fits generalized additive models)
- lm (linear model)



盒形圖 (Boxplot)

- Plotting with a categorical explanatory variable: boxplot.
- Categorical variables are factors with two or more levels.
- Boxplot
  - The horizontal line shows the median.
  - The bottom and top of the box show the 25th and 75th percentiles.
  - The vertical dashed lines are called the "whiskers".
  - Either maximum or 1.5 times the IQR.
- boxplot not only show the location and spread of data but also indicate skewness.

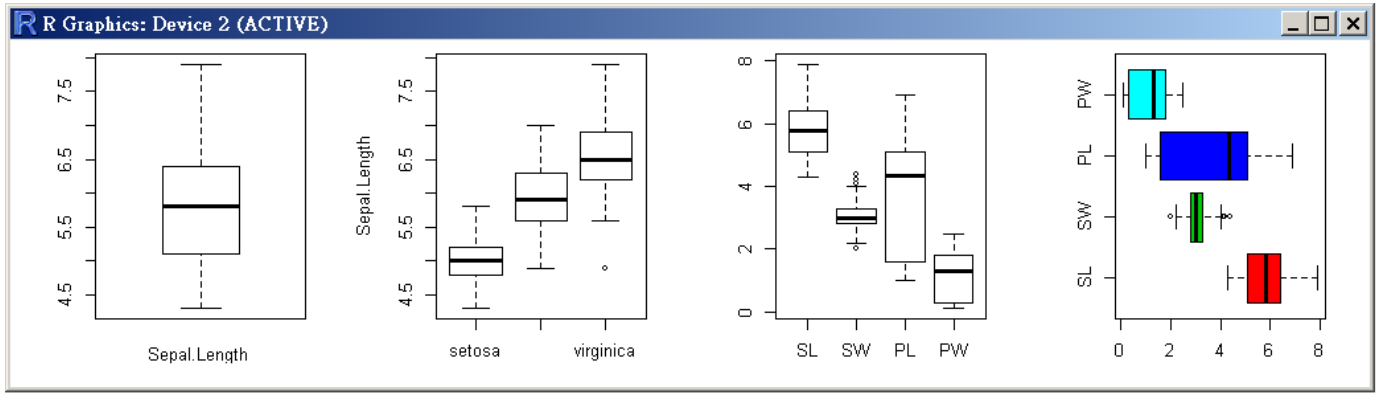


- Is a factor significant?
- Does the location differ between subgroups?
- Does the variation differ between subgroups?
- Are there any outliers?

boxplot(x, ...)

```
## S3 method for class 'formula'
boxplot(formula, data = NULL, ..., subset, na.action = NULL)
## Default S3 method:
boxplot(x, ..., range = 1.5, width = NULL, varwidth = FALSE,
        notch = FALSE, outline = TRUE, names, plot = TRUE,
        border = par("fg"), col = NULL, log = "",
        pars = list(boxwex = 0.8, staplewex = 0.5, outwex = 0.5),
        horizontal = FALSE, add = FALSE, at = NULL)
```

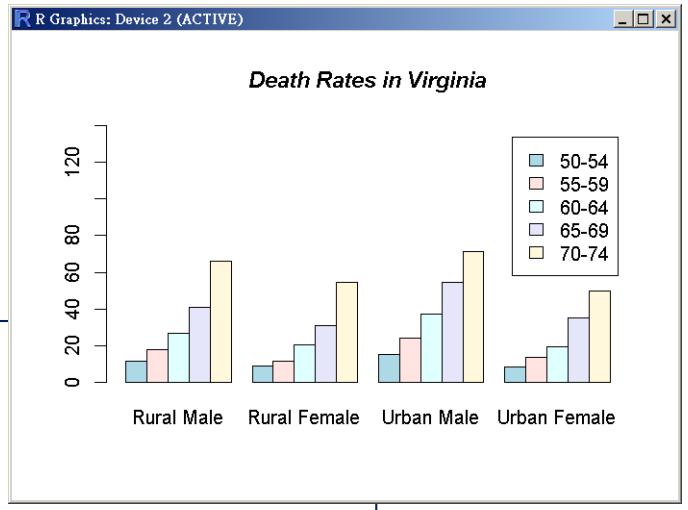
```
> par(mfrow=c(1,4))
> names(iris)
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
> names(iris) <- c("SL", "SW", "PL", "PW", "Species")
> boxplot(Sepal.Length, xlab="Sepal.Length")
> boxplot(Sepal.Length~Species, ylab="Sepal.Length")
> boxplot(iris[,which(sapply(iris, is.numeric))])
> boxplot(iris[,which(sapply(iris, is.numeric))], horizontal=T, col=2:8)
```



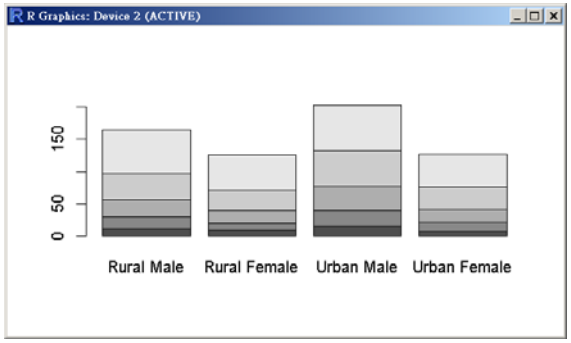
長條圖 (barplot)

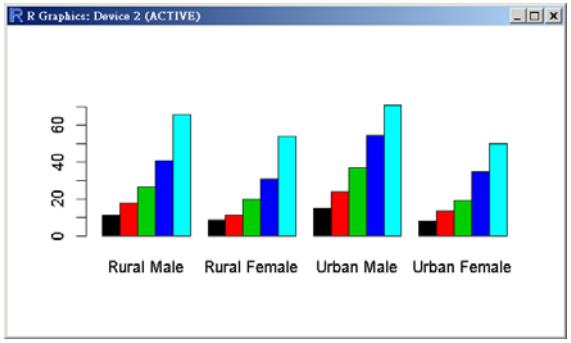
Death Rates in Virginia (1940): The death rates are measured per 1000 population per year. They are cross-classified by age group (rows) and population group (columns). The age groups are: 50–54, 55–59, 60–64, 65–69, 70–74 and the population groups are Rural/Male, Rural/Female, Urban/Male and Urban/Female.

```
> VADeaths
      Rural Male Rural Female Urban Male Urban Female
50-54       11.7          8.7       15.4          8.4
55-59       18.1         11.7       24.3         13.6
60-64       26.9         20.3       37.0         19.3
65-69       41.0         30.9       54.6         35.1
70-74       66.0         54.3       71.1         50.0
```



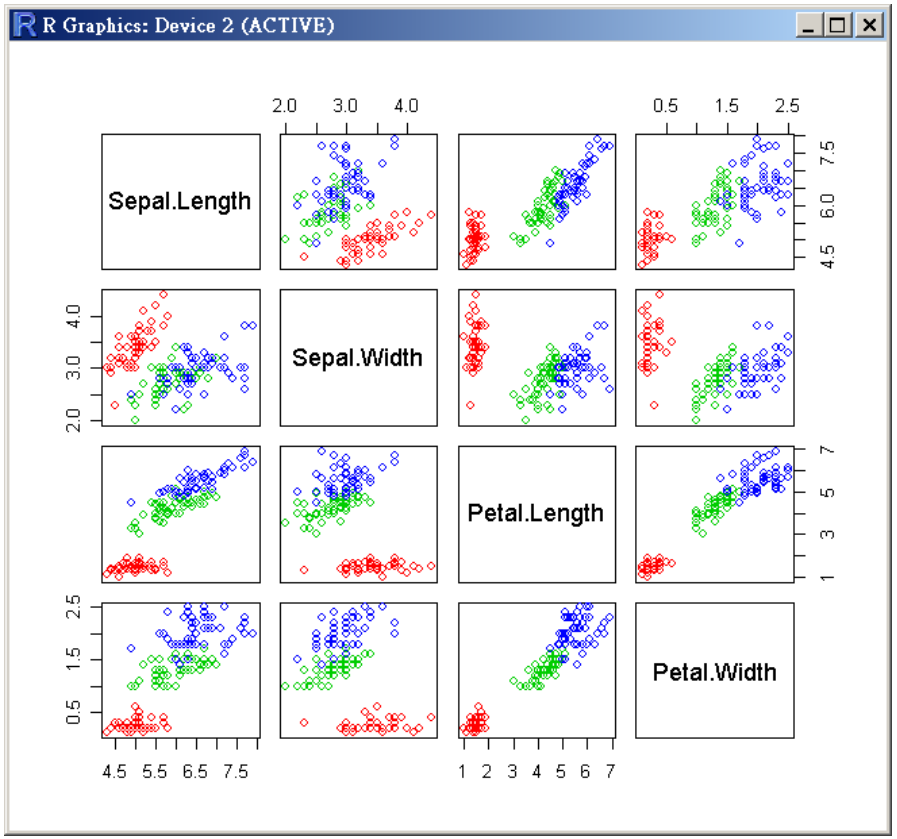

```
barplot(VADeaths)
barplot(VADeaths, beside = TRUE, col=1:5)

barplot(VADeaths, beside = TRUE,
        col = c("lightblue", "mistyrose", "lightcyan",
                "lavender", "cornsilk"),
        legend = rownames(VADeaths), ylim = c(0, 140))
title(main = "Death Rates in Virginia", font.main = 4)
```

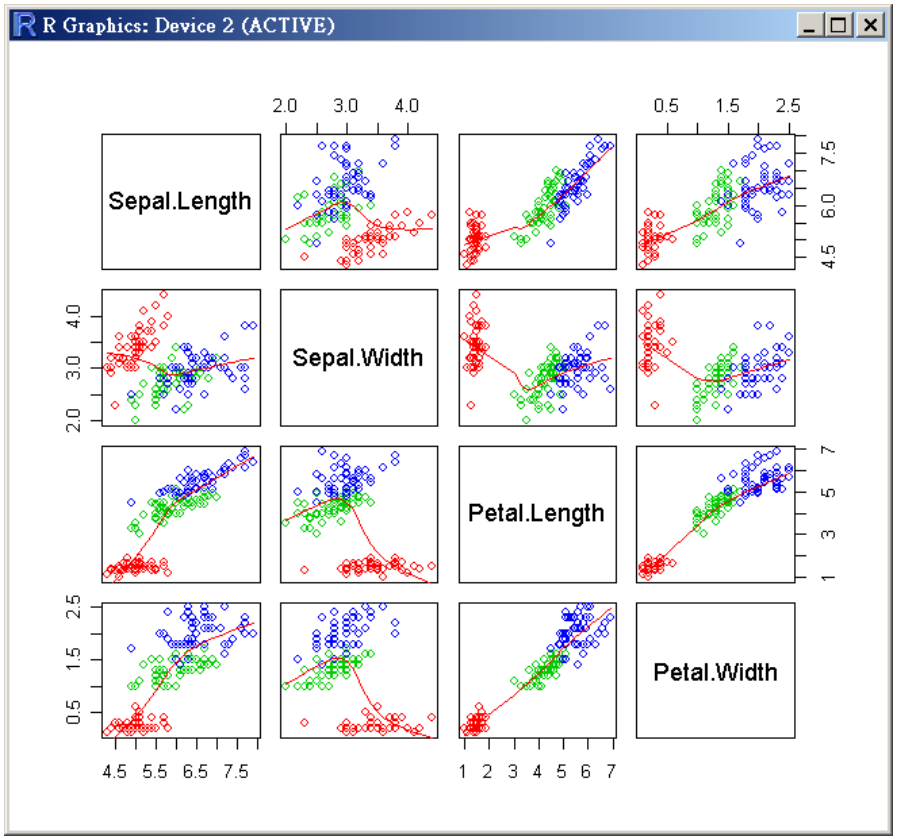




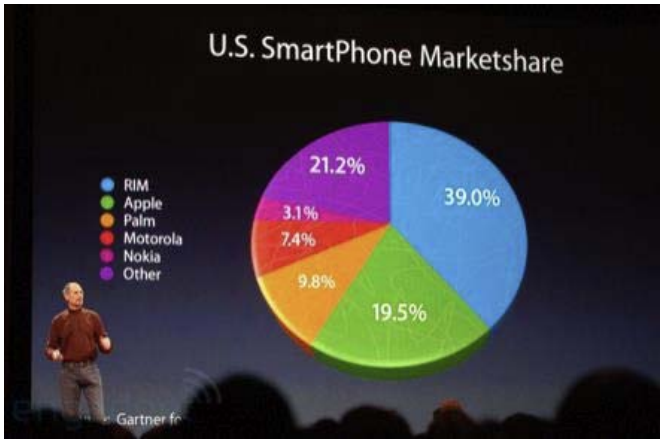
散佈圖矩陣 (Scatterplot Matrices)

```
> pairs(iris[,1:4], col=as.integer(iris[,5])+1)
> pairs(iris[,1:4], col=as.integer(iris[,5])+1, panel=panel.smooth)
```

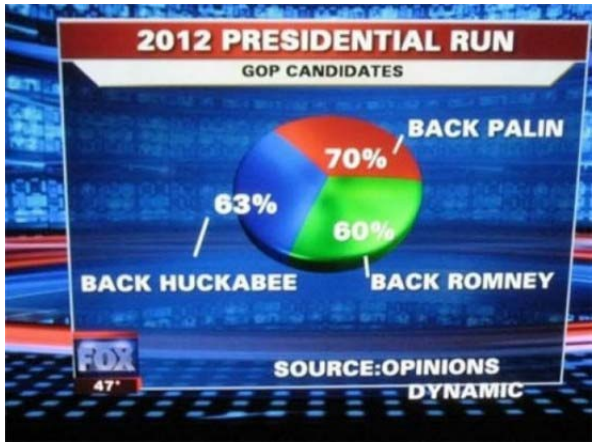


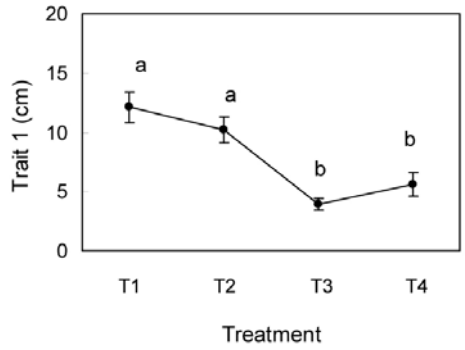


圖表的誤用

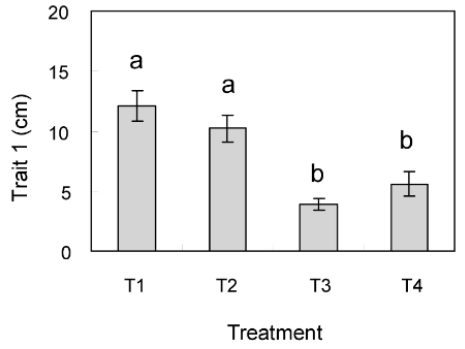


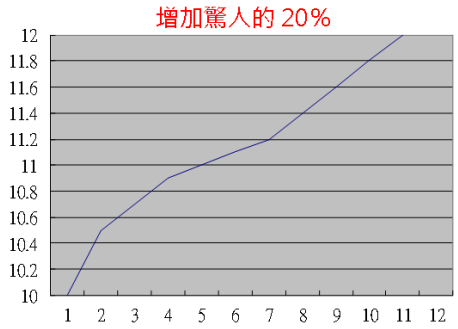
Source: https://www.managertoday.com.tw/articles/view/51480

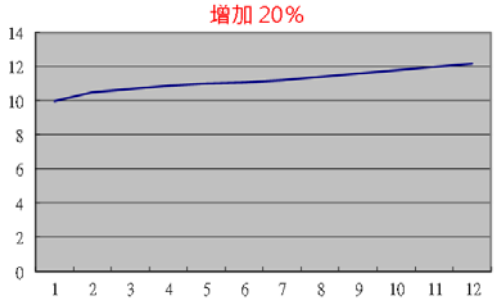




Source: http://ir.tari.gov.tw:8080/bitstream/345210000/3094/1/journal_arc_60-1-6.pdf





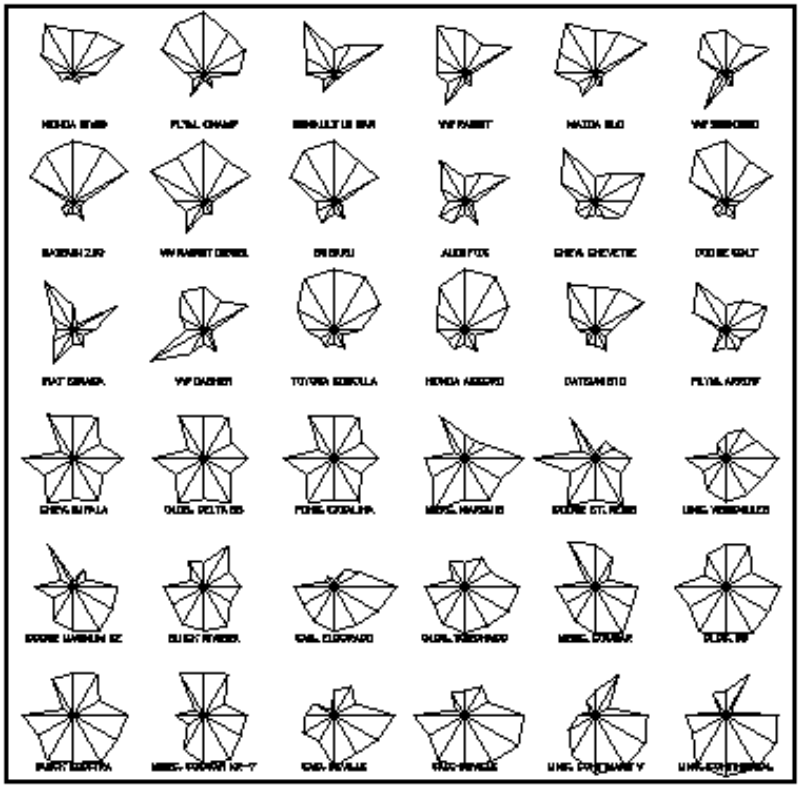


Misleading Graphs: Real Life Examples
http://www.statisticshowto.com/misleading-graphs/
The top ten worst graphs
https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/
Bad Infographics: 11 Mistakes You Never Want to Make
http://blog.visme.co/bad-infographics/
13 Graphs That Are Clearly Lying
https://www.buzzfeed.com/katienotopoulos/graphs-that-lied-to-us?utm_term=.qsnBZa6Qa#.xePkLjDaj
11 Most Useless And Misleading Infographics On The Internet
https://io9.gizmodo.com/11-most-useless-and-misleading-infographics-on-the-inte-1688239674
The most misleading charts of 2015, fixed
https://qz.com/580859/the-most-misleading-charts-of-2015-fixed/
Misleading graph
https://en.wikipedia.org/wiki/Misleading_graph

Star Plot (Chambers 1983)

- The star plot consists of a sequence of equi-angular spokes, called radii, with each spoke representing one of the variables.
  - The data length of a spoke is proportional to the magnitude of the variable for the data point.
  - A line is drawn connecting the data values for each spoke.
- Typically, star plots are generated in a multi-plot format with many stars on each page and each star representing one observation.

Star plot of Automobile Data



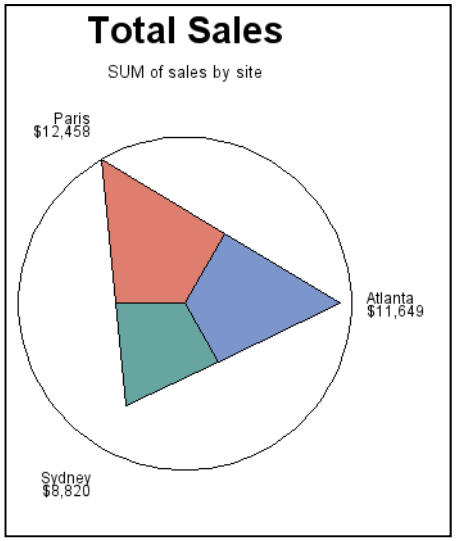
- Each star represents one car model.
- Each ray in the star is proportional to one variable.
- The dominant pattern is that the star symbols in the top rows have long rays on the top (good price and performance) and short rays on the bottom (small in size variables), but the reverse is generally true for the heaviest models in the bottom rows.
- The primary weakness of star plot is that their effectiveness is limited to data sets with less than a few hundred points.

Variants of Star Plot

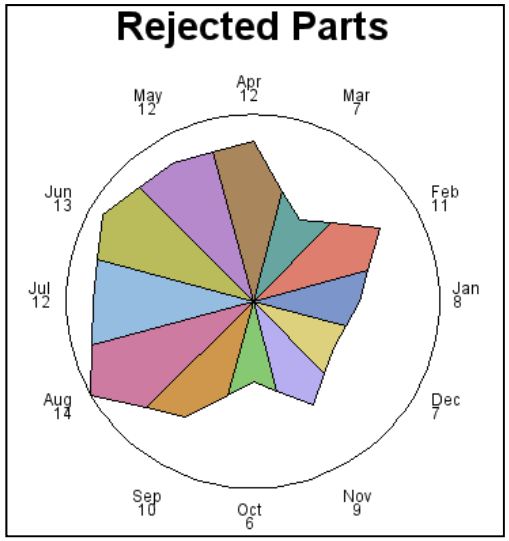
The star plot can show:

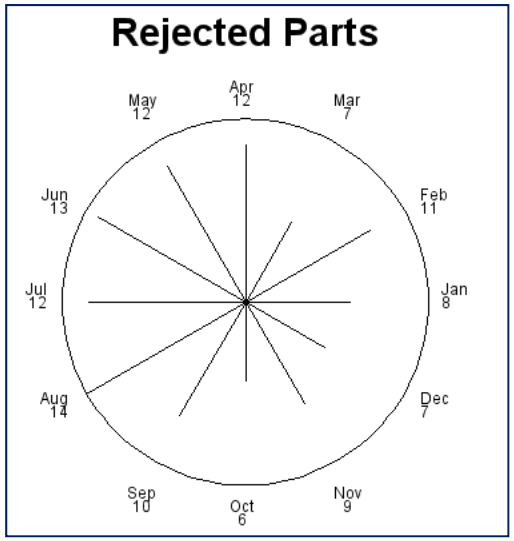
- What variables are dominant for a given observation?
- Which observations are most similar, i.e., are there clusters of observations?
- Are there outliers?

Specifying the Sum Statistic in a Star Chart



Charting a Discrete Numeric Variable in a Star Chart

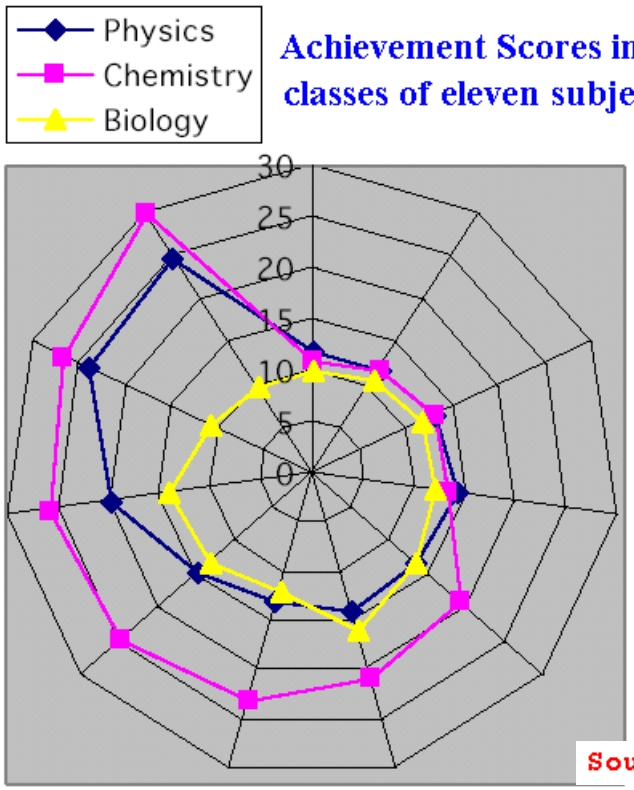




Figures Source: SAS/GRAPH® 9.2 Reference, 2nd Edition

Radar Plot

- The idea of a radar plot is similar to that of the star plot.
- In a radar plot, the value of the measurement is also represented by radii stretching out from the center of a circle.
- However, here each radius stands for a subject instead of a variable.
- The subjects' response on each variable is displayed by points of different shapes, colors, or both.



- The graph shows the frequencies of data series relative to one another.
- Apparently, the biology scores are the lowest and are not correlated to neither physics nor chemistry scores.
- However, physics and chemistry scores are good predictors to each other.

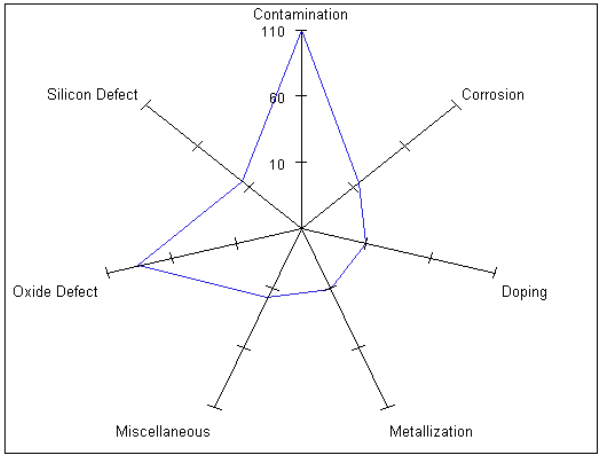
This approach suffers the same shortcoming as the star graph. When there are too many variables and subjects, the data pattern will be concealed.

Source: ATKOSoft: Survey on Visualisation methods and software tools

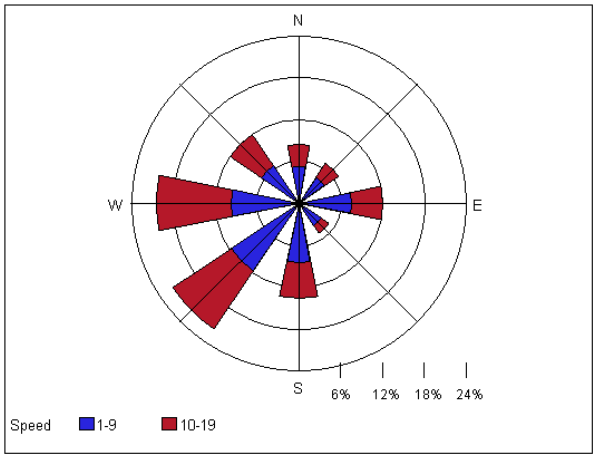
Variants of Radar Plot

Figures Source: SAS/GRAPH® 9.2 Reference, 2nd Edition

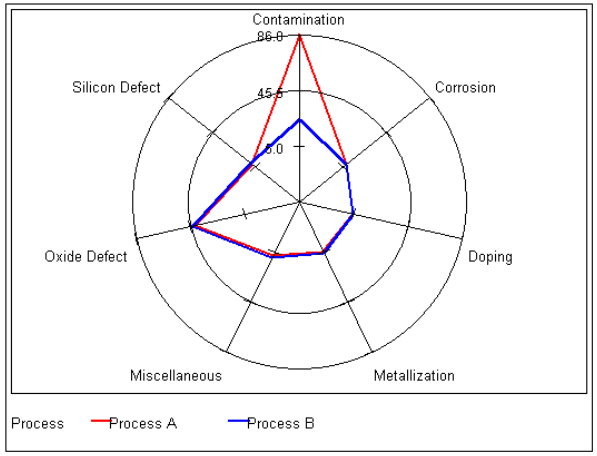
Producing a Basic Radar Chart



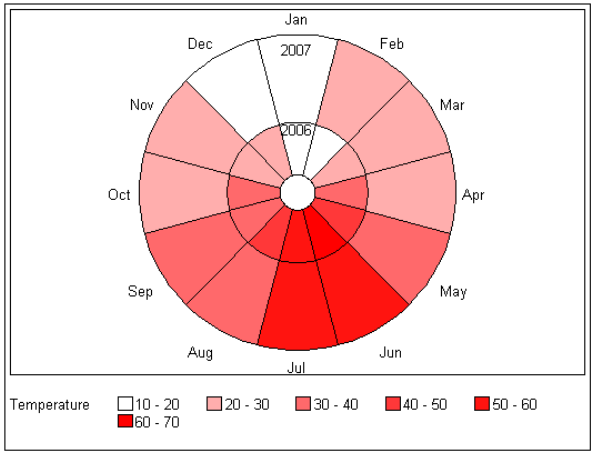
Creating a Windrose Chart



Modifying the Appearance of Radar Charts

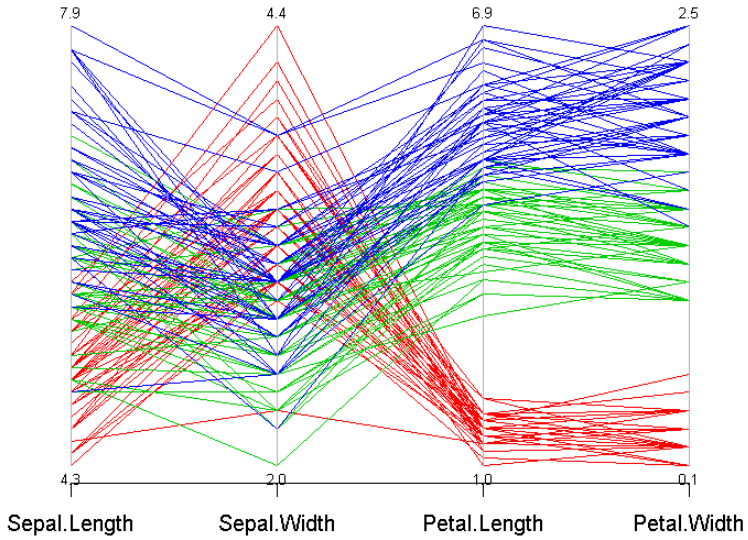


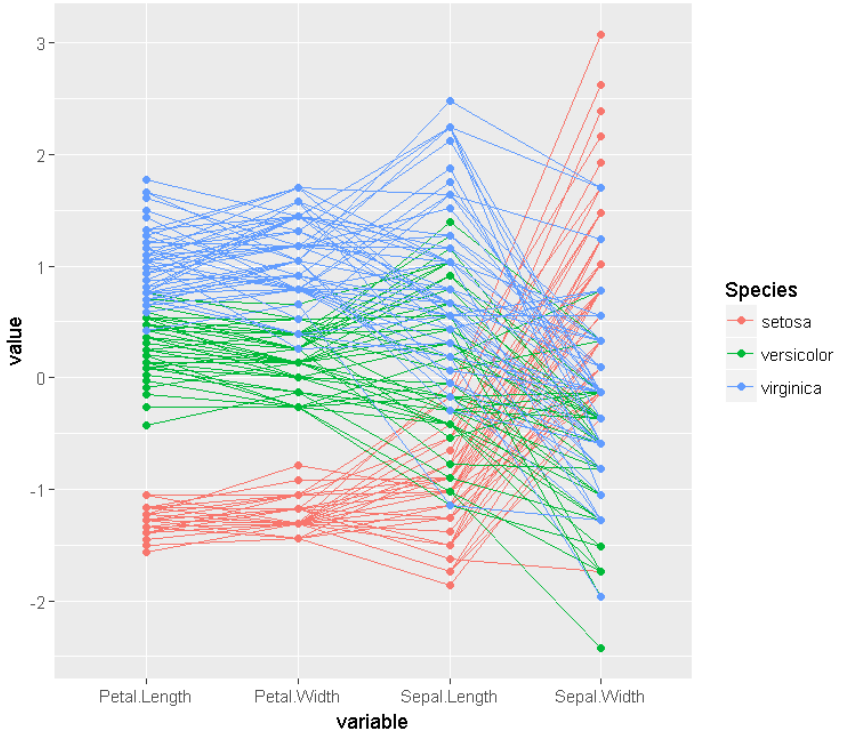
Creating a Calendar Chart



Parallel Coordinates Plot

```
> library(MASS)
> parcoord(iris[,1:4], col=as.integer(iris[,5])+1, var.label = T)
> library(GGally) # Extension to 'ggplot2'
> ggparcoord(data = iris, columns = 1:4, groupColumn = 5)
> ggparcoord(data = iris, columns = 1:4, groupColumn = 5, boxplot = T)
> ggparcoord(data = iris, columns = 1:4, groupColumn = 5, order = "anyClass",
+ showPoints = TRUE)
```



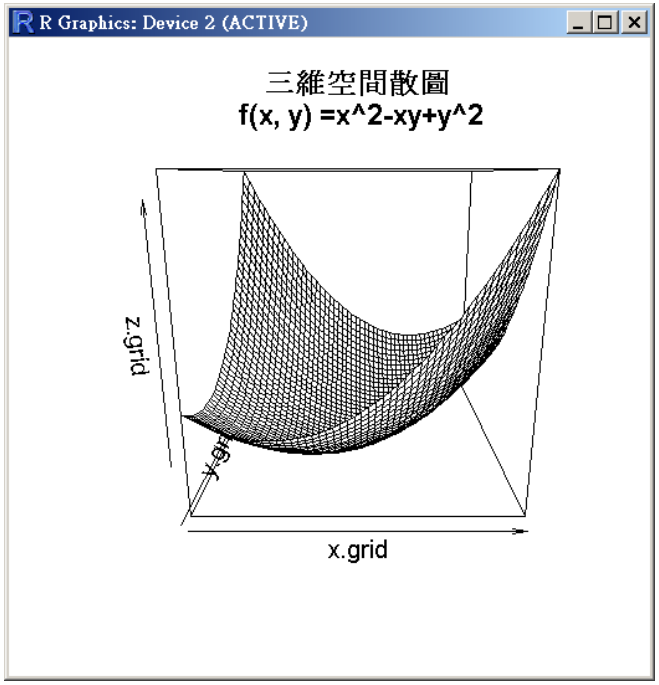



```
ggparcoord(data, columns = 1:ncol(data),
  groupColumn = NULL,
  scale = "std", scaleSummary = "mean", centerObsID = 1,
  missing = "exclude", order = columns, showPoints = FALSE,
  splineFactor = FALSE, alphaLines = 1, boxplot = FALSE,
  shadeBox = NULL, mapping = NULL, title = "")
```

3D透視圖: persp

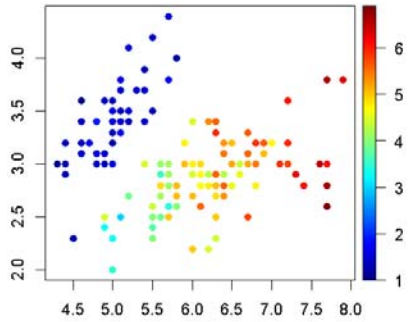
- 雙變數函數在三維空間的散佈圖
- NOTE: x-y平面上，格點數目愈多，散佈圖愈密集。
- plot f(x, y) =x^2-xy+y^2。

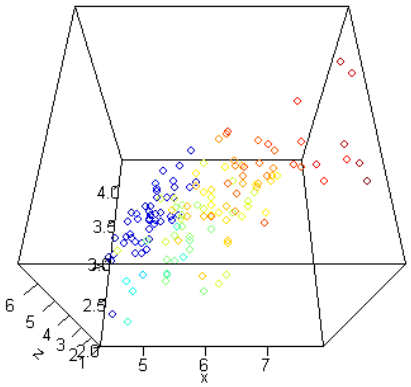
```
> ploy <- function(x, y){x^2-x*y+y^2}
> x.grid <- seq(-3, 3, length=50)
> y.grid <- seq(-3, 3, length=50)
> z.grid <- outer(x.grid, y.grid, FUN=ploy)
> ploy.title <- paste("三維空間散圖\n", "f(x, y) =x^2-xy+y^2")
> persp(x.grid, y.grid, z.grid, main= ploy.title)
```

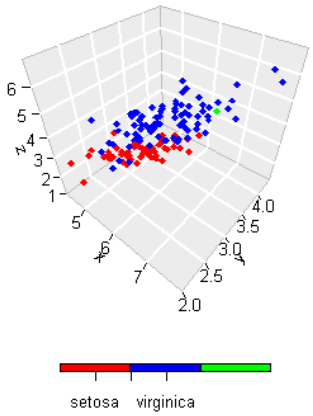


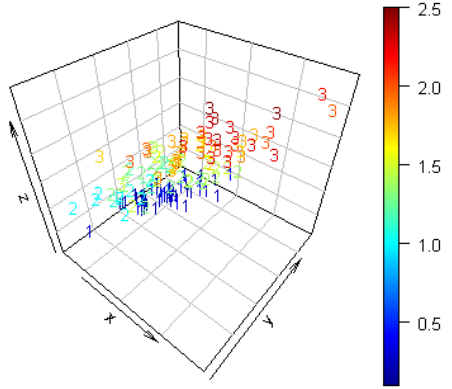
plot3D package: 3D散佈圖

```
par(mfrow = c(1,3), mai = c(0.3, 0.3, 0.2, 0.3))
# grey background with white grid lines
scatter3D(x, y, z, bty = "f", colkey = FALSE, ticktype = "detailed",
          theta=0, phi=60)
# theta: the azimuthal direction, phi: the co-latitude.
scatter3D(x, y, z, bty = "g", pch = 18,
          col.var = as.integer(s),
          col = c("red", "blue", "green"),
          pch = 18, ticktype = "detailed",
          colkey = list(at = c(2, 3, 4), side = 1,
                        addlines = TRUE, length = 0.5, width = 0.5,
                        labels = c("setosa", "versicolor", "virginica")))
text3D(x, y, z, labels = as.integer(s), colvar = w, bty = "b2")
# "b2": back panels and grid lines are visible

# hist3D, text3D
# scatter2D(x, y, colvar = z, pch = 16)
```







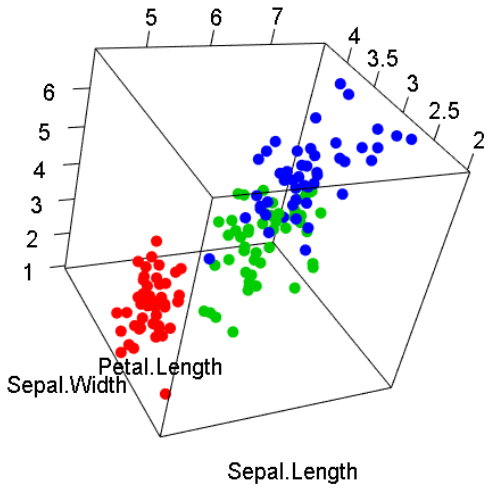


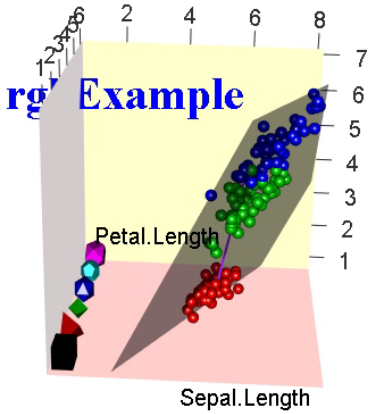
http://www.sthda.com/english/wiki/impressive-package-for-3d-and-4d-graph-r-software-and-data-visualization

rgl

3D visualization device system (OpenGL)

```
> library("rgl")
> open3d()
> plot3d(iris[,1:3], col=as.integer(iris[,5])+1, type ="p", size=10)
```

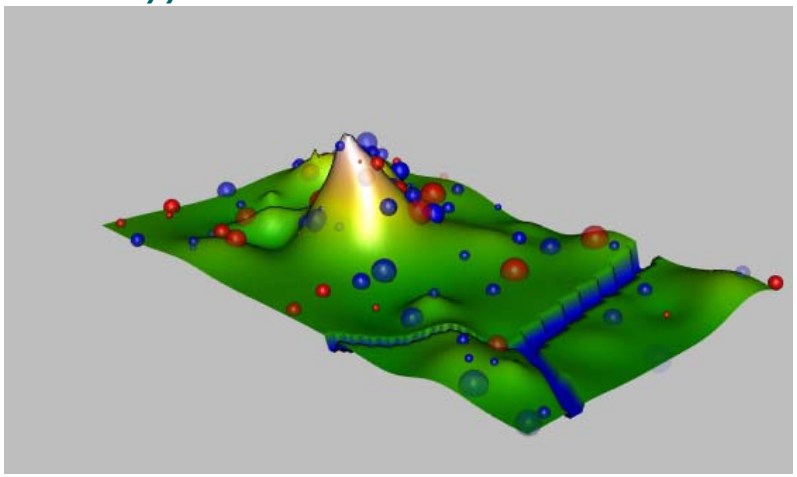





```
> plot3d(iris[,1:3], col=as.integer(iris[,5])+1, type ="s",
+        radius=0.15)
> bbox3d(color=c("red", "black"), emission="gray",
+        specular="yellow", shininess=5, alpha=0.8, nticks = 3)
> aspect3d(1,1,1)
>
> lines3d(iris[c(1, 150), 1:3], col="purple", lwd=2)
> # points3d, lines3d,segments3d,triangles3d, quads3d.
>
> shapes <- list(cube3d(), tetrahedron3d(), octahedron3d(),
+           icosahedron3d(), dodecahedron3d(), cuboctahedron3d())
> shapelist3d(shapes, x=1, y=1:6, z=1, size=0.3, col=1:6)
> aspect3d(1,1,1)
>
> texts3d(x=2, y=6, z=6, texts="rgl Example", font=2,
+         color="blue", cex=2, family="serif")
>
> # Show regression plane with z as dependent variable
> fit <- lm(iris[,3] ~ iris[,1] + iris[,2])
> coefs <- coef(fit)
> planes3d(a=coefs[2], b=coefs[3], c=-1, d= coefs["(Intercept)"],
+          alpha = 0.5)
> # planes3d draws planes using ax + by + cz + d = 0.
>
> play3d(spin3d(axis = c(0, 0, 1), rpm = 20), duration = 4)
```

範例: rgl

```
> terrain <- as.matrix(read.table("terrain_data.txt", header=F))
> dim(terrain)
[1] 100 100
> animal <- read.table("animal_data.txt", header=T)
> dim(animal)
[1] 100   5
> attach(animal)
> head(animal, 3)
      loc.x     loc.y number sex    index
1  1.421804 0.1536418      4   1 7.571144
2 86.589918 0.7205304      3   1 6.855152
3 58.427946 2.1297322      2   0 7.526096
> # Define colors for terrain
> # see ? surface3d for how to assign colors
> terrain.scale <- floor((terrain - min(terrain))/(max(terrain) - min(terrain))*99)+1
> terrain.color <- terrain.colors(100)[terrain.scale]
> terrain.color[terrain==0] <- rgb(0, 0, 1) # set color for river
> open3d()
> clear3d("all")
> bg3d(col="gray") # setup background
> light3d() # setup head-light
> surface3d(1:100, seq(1,60,length=100), terrain, col=terrain.color, back="lines")
> sex.col <- ifelse(sex==0, rgb(0, 0, 1), rgb(1, 0, 0)) # males: blue, females: red
>
> z <- terrain[cbind(ceiling(loc.x), ceiling(loc.y*10/6))]
> alpha.index <- (index-min(index))/(max(index)-min(index))
> spheres3d(loc.x, loc.y, z + 0.5,
+           radius=0.3*number, col=sex.col,
+           alpha=alpha.index)
> detach(animal)
> play3d(spin3d(), duration=10)
```



```
# this example is modified
from rgl.demo.abundance()
# see rgl.demo.abundance
```

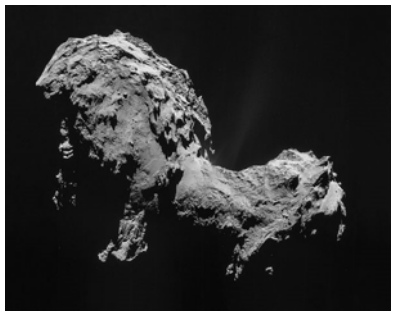
範例: rgl, explore a comet

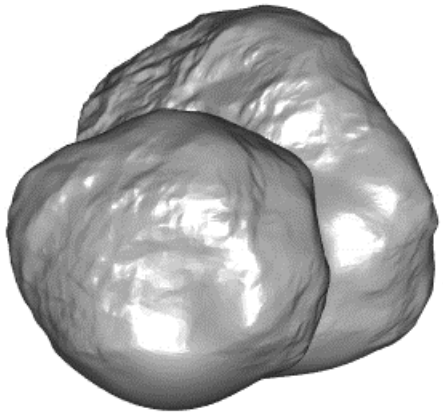
Explore a comet with R's "rgl" package

December 24, 2014

http://blog.revolutionanalytics.com/2014/12/explore-a-comet-with-rs-rgl-package.html

"Last month, the Philae lander touched down on comet Churyumov–Gerasimenko. In the process, the lander and the orbiting Rosetta probe captured detailed data on the geometry of the comet, which the ESA published as a shape file. ..."





https://en.wikipedia.org/wiki/67P/Churyumov%E2%80%93Gerasimenko

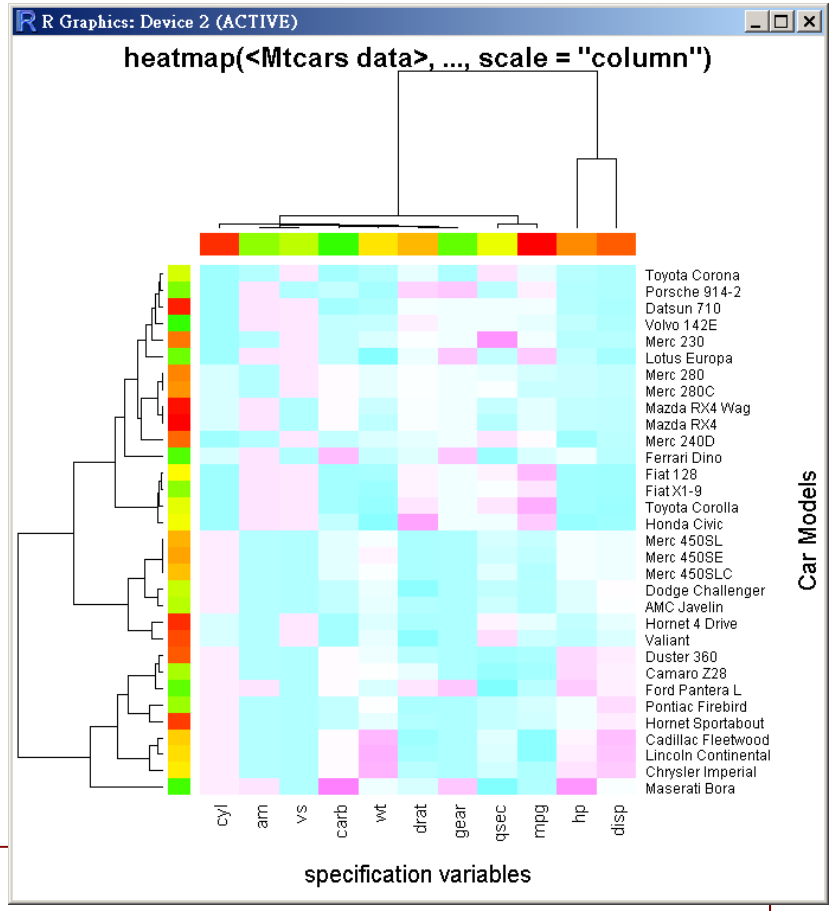
```
> open3d()
> # comet <- readOBJ(url("http://sci.esa.int/science-e/www/object/doc.cfm?fobjectid=54726"))
> comet <- readOBJ("ESA_Rosetta_OSIRIS_67P_SHAP2P.obj")
> class(comet)
[1] "mesh3d"  "shape3d"
> str(comet)
List of 6
 $ vb           : num [1:4, 1:31456] -0.394 0.402 0.443 1 -0.163 ...
 $ it           : num [1:3, 1:62908] 14327 6959 18747 8258 15598 ...
 $ primitivetype: chr "triangle"
 $ material     : NULL
 $ normals      : NULL
 $ texcoords    : NULL
 - attr(*, "class")= chr [1:2] "mesh3d" "shape3d"
> shade3d(comet, col="gray")
```

```
# it: indices for triangular faces
# ib: indices for quad faces
# vb: matrix of vertices: 4xn matrix (rows
x, y, z, h) or equivalent vector, where h
indicates scaling of each plotted quad
```

熱圖 (Heatmap)

- **heatmap {stats}**: A heat map is a false color image (basically **image(t(x))**) with a dendrogram added to the left side and to the top.
- A deafult dendrograms are computed as **as.dendrogram(hclustfun(distfun(X)))**

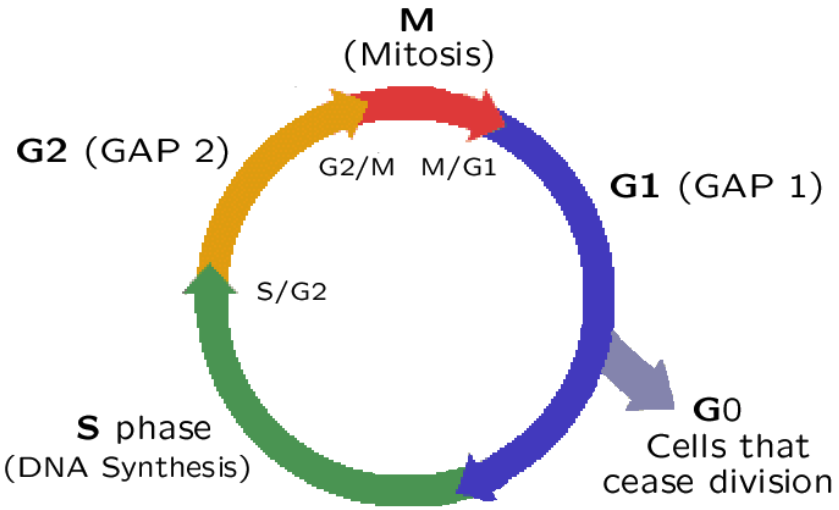
```
> mtcars
                   mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
```

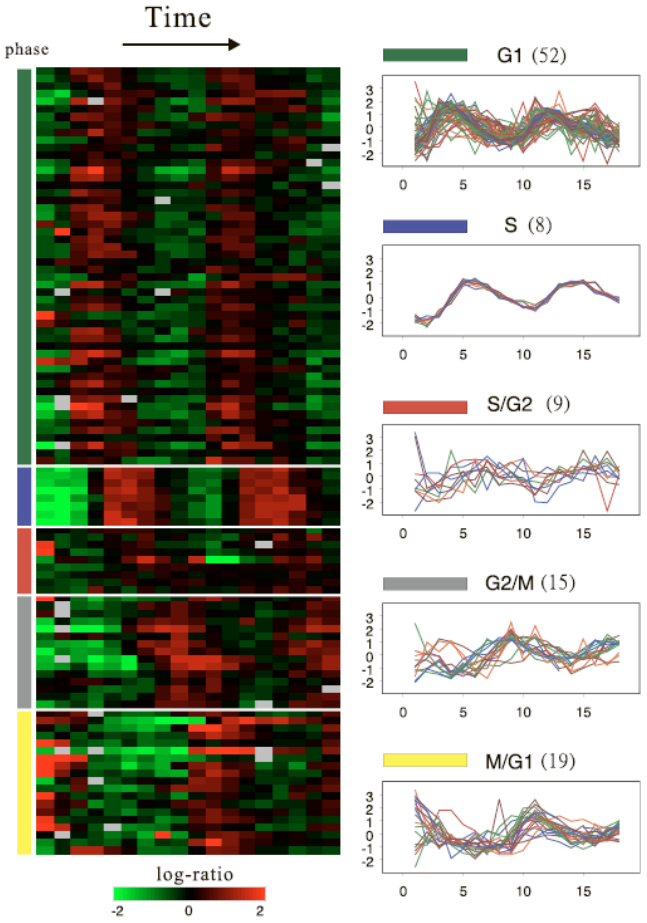



```
x  <- as.matrix(mtcars)
rc <- rainbow(nrow(x), start = 0, end = .3)
cc <- rainbow(ncol(x), start = 0, end = .3)
hv <- heatmap(x, col = cm.colors(256), scale = "column",
              RowSideColors = rc, ColSideColors = cc, margins = c(5,10),
              xlab = "specification variables", ylab =  "Car Models",
              main = "heatmap(<Mtcars data>, ..., scale = \"column\")")
```

課堂練習: Microarray Data

- 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.)



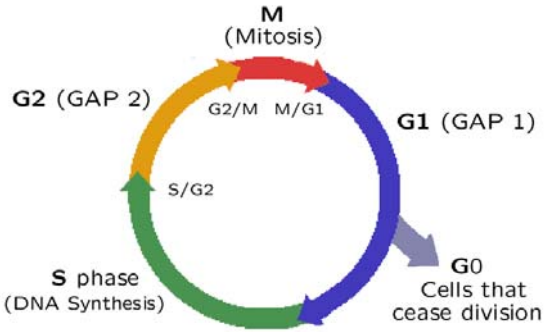


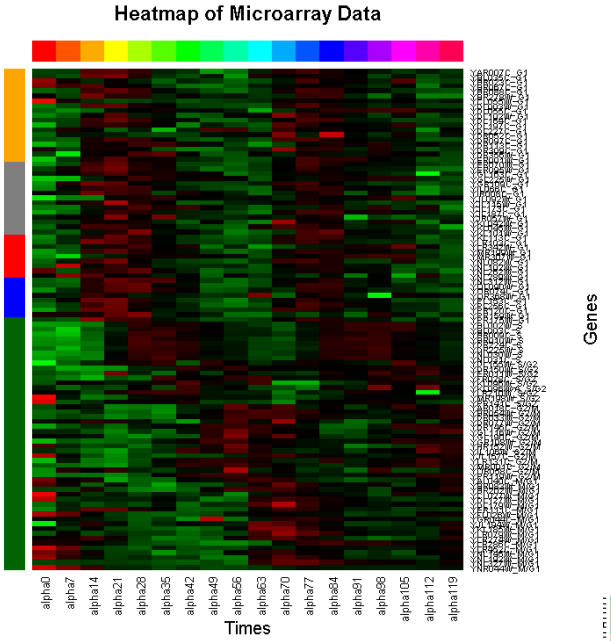
See also: Using R to draw a Heatmap from Microarray Data
http://www2.warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/heatmap/

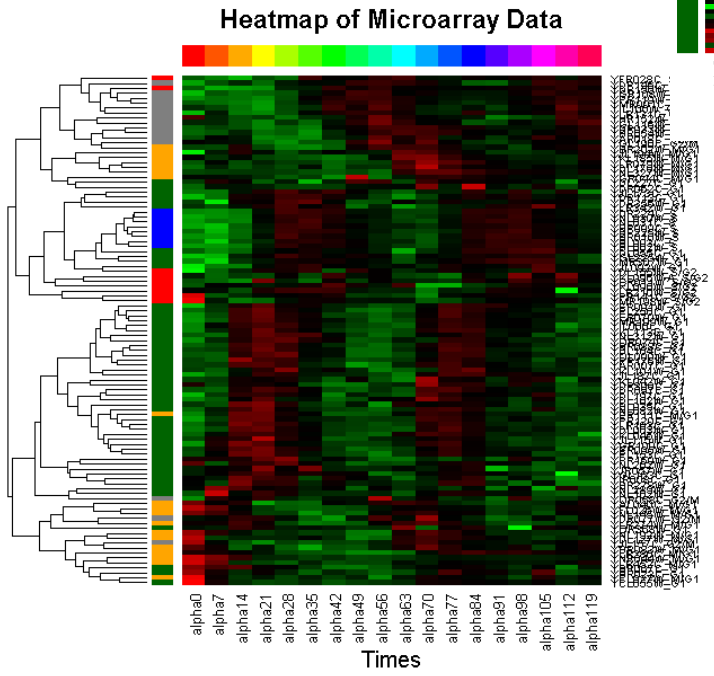
課堂練習: Microarray Data

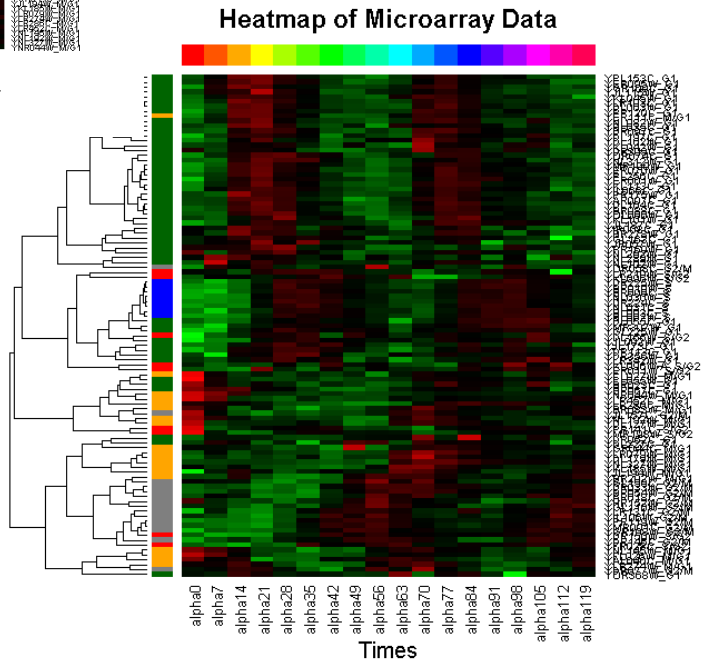
```
# install.packages("fields")
library(fields)
gbr <- two.colors(start="green", middle="black", end="red")
cell.raw <- read.table("trad_alpha103.txt", row.names=1, header=T)
cell.data <- t(scale(t(cell.raw[,2:19]), center=T, scale=T))
n <- nrow(cell.data)
p <- ncol(cell.data)
gene.phase <- cell.raw[,1]
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



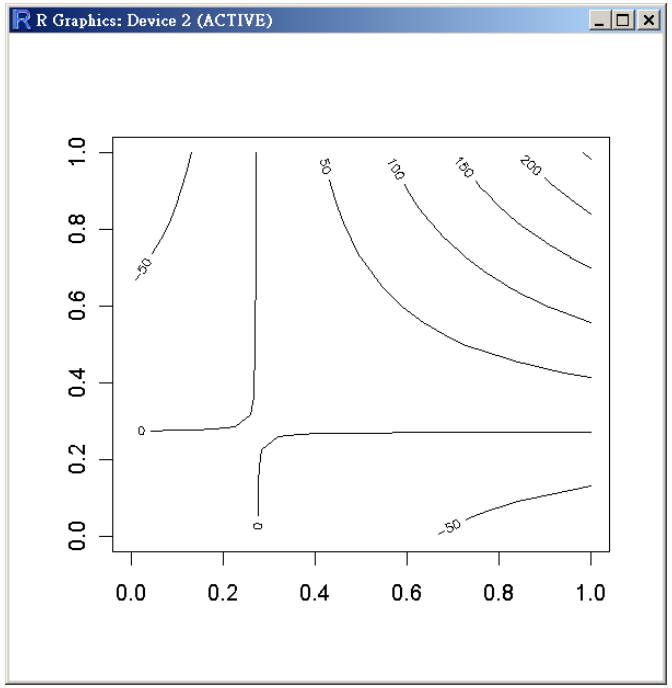




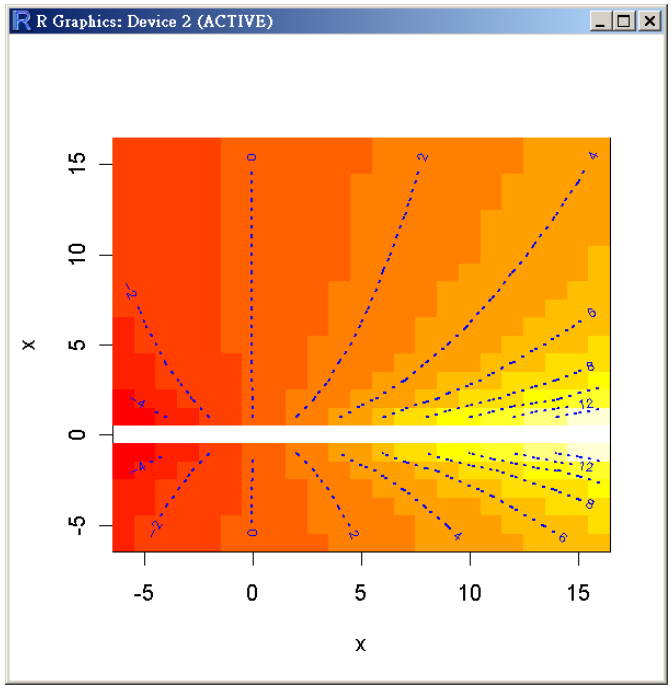


等高線圖 (Contours)

```
> x <- -6:16; length(x)
> my.data <- outer(x, x);dim(my.data)
> my.data
> contour(my.data, method = "edge")
```

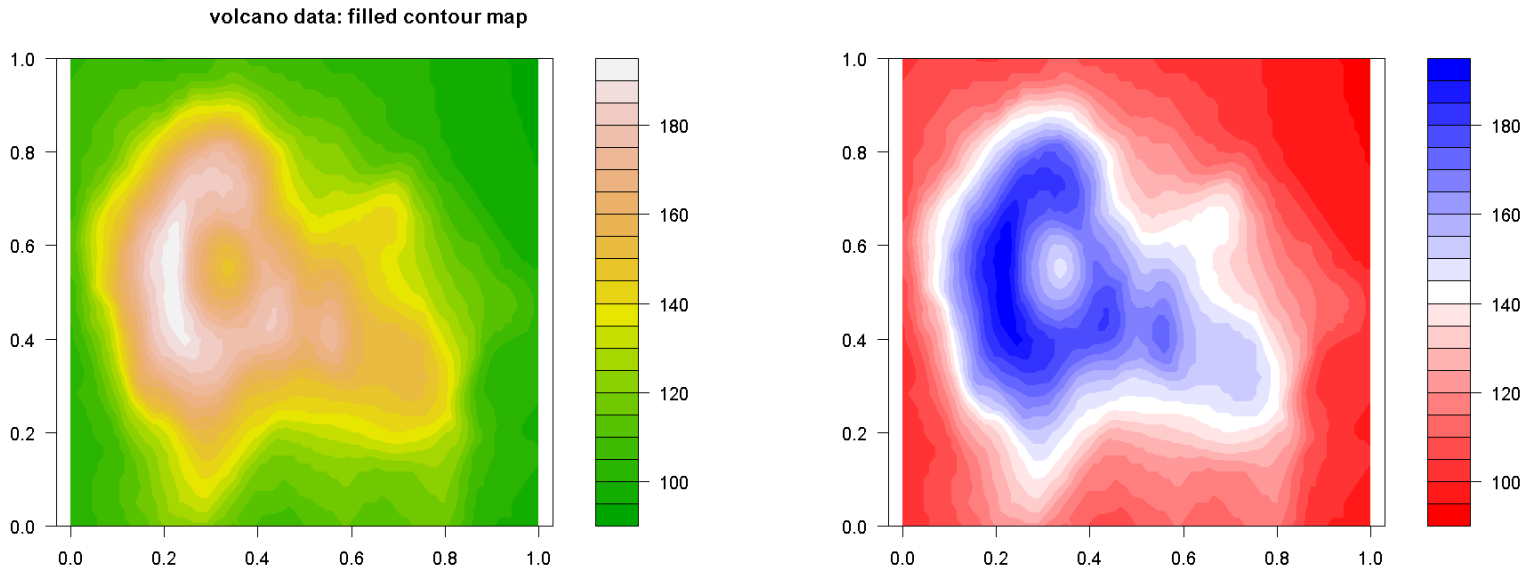



```
> image(x, x, z)
> contour(x, x, z, col = "blue", add = TRUE,
          method = "edge", lwd=1.5, lty=3)
```



Filled Contours

```
> filled.contour(volcano, color.palette = terrain.colors, asp = 1)
> title(main = "volcano data: filled contour map")
> # asp: the y/x aspect ratio.
> filled.contour(volcano, color.palette = colorRampPalette(c("red", "white",
"blue")), asp = 1)
```



讀取外部影像檔案

```
> install.packages(c("tiff", "jpeg", "png", "fftwtools"),
repos="http://cran.csie.ntu.edu.tw")
> library(EBImage) # (Repositories: BioC Software)
> Transformers <- readImage("Transformers07.jpg")
> (dims <- dim(Transformers))
[1] 300 421   3
> Transformers
Image
  colorMode    : Color
  storage.mode : double
  dim          : 300 421 3
  frames.total : 3
  frames.render: 1

imageData(object)[1:5,1:6,1]
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    0    0    0    0    0
[2,]    0    0    0    0    0    0
[3,]    0    0    0    0    0    0
[4,]    0    0    0    0    0    0
[5,]    0    0    0    0    0    0
> plot(c(0, dims[1]), c(0, dims[2]), type='n',
+ xlab="", ylab="")
> rasterImage(Transformers, 0, 0, dims[1], dims[2])
```

```
> source("https://bioconductor.org/biocLite.R")
> biocLite("EBImage")
```


```
> #install.packages("jpeg")
> library(jpeg)
> Transformers <- readJPEG("Transformers07.jpg")
```

https://en.wikipedia.org/wiki/Transformers_(film)

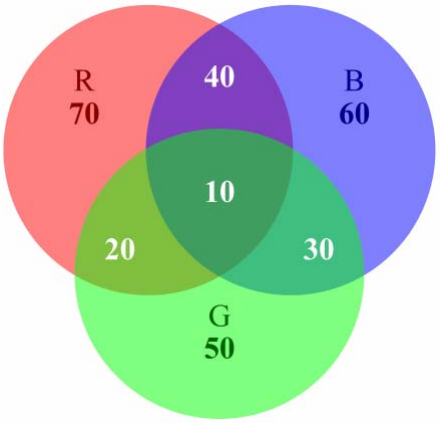
彩色影像轉成灰階

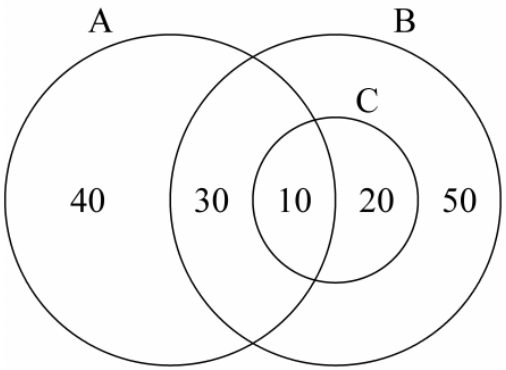
```
> Transformers.f <- Image(flip(Transformers))
> # convert RGB to grayscale
> rgb.weight <- c(0.2989, 0.587, 0.114)
> Transformers.gray <- rgb.weight[1] * imageData(Transformers.f)[,,1] +
+                      rgb.weight[2] * imageData(Transformers.f)[,,2] +
+                      rgb.weight[3] * imageData(Transformers.f)[,,3]
> dim(Transformers.gray)
[1] 300 421
> Transformers.gray[1:5, 1:5]
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    0    0    0
[2,]    0    0    0    0    0
[3,]    0    0    0    0    0
[4,]    0    0    0    0    0
[5,]    0    0    0    0    0
> par(mfrow=c(1,2), mai=c(0.1, 0.1, 0.1, 0.1))
> image(Transformers.gray, col = grey(
+ seq(0, 1, length = 256)), xaxt="n", yaxt="n")
> image(Transformers.gray, col = rainbow(256),
+ xaxt="n", yaxt="n")
```

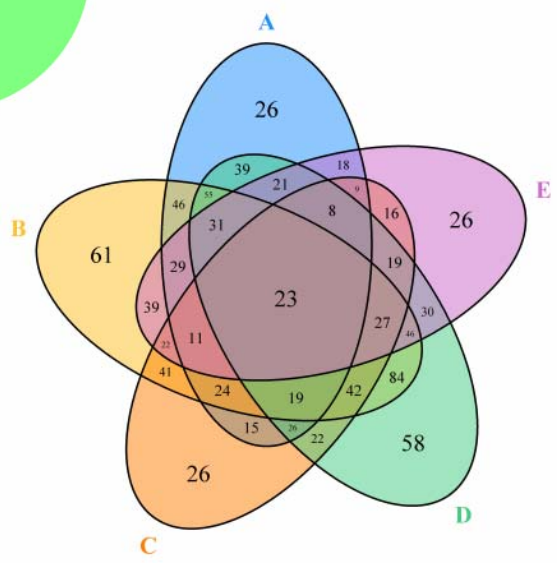
Converting RGB to grayscale/intensity
http://stackoverflow.com/questions/687261/converting-rgb-to-grayscale-intensity

Venn Diagrams

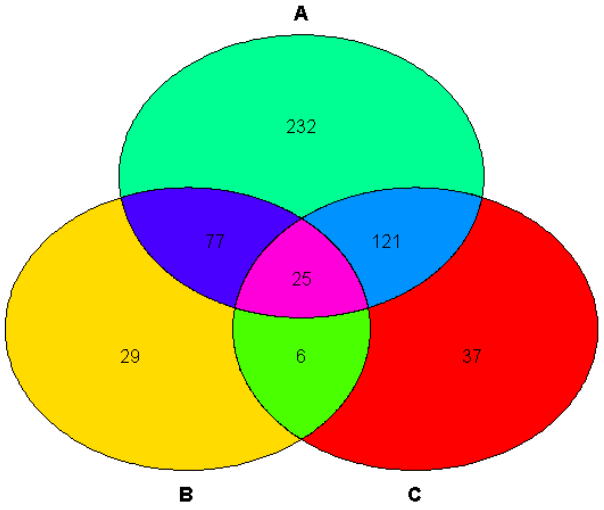
Package **VennDiagram**

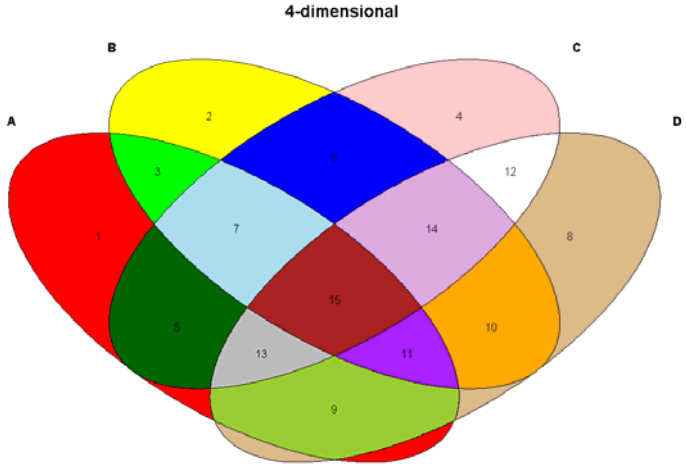






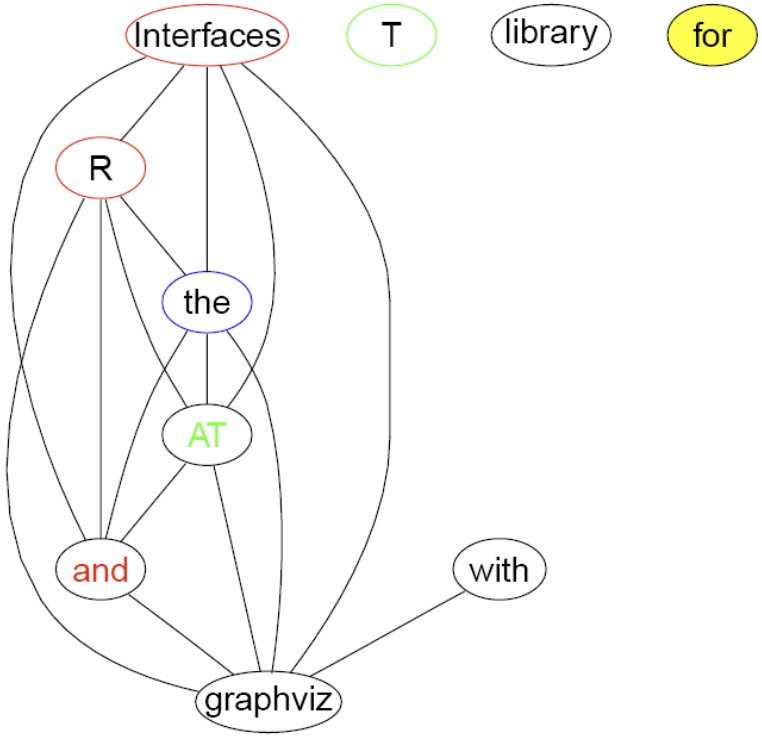
Package **colorfulVennPlot**





Networks: Rgraphviz, igraph

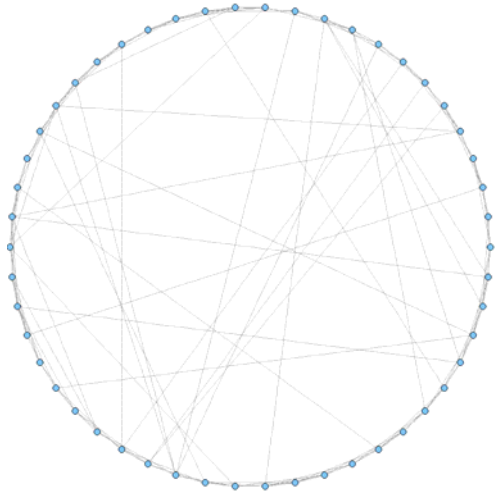
Rgraphviz: Interfaces R with the AT and T graphviz library for plotting R graph objects from the graph package.

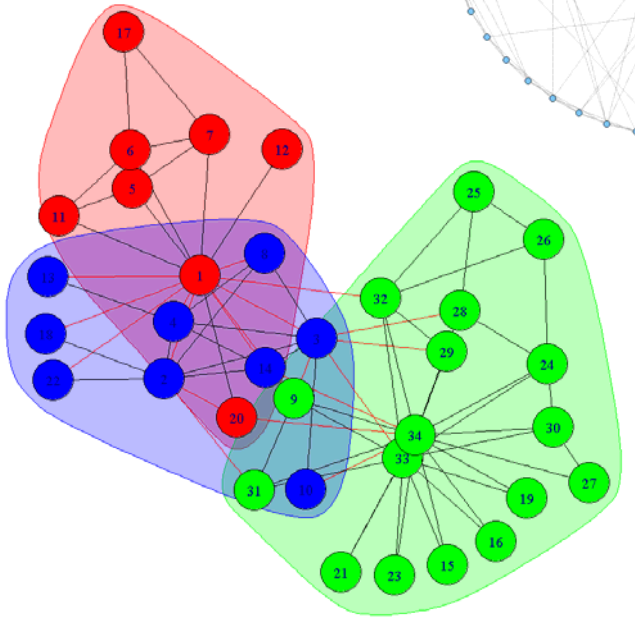


The network analysis package

```
demo(package="igraph")
demo(package="igraph", community)
demo(package="igraph", smallworld)
```

http://igraph.org





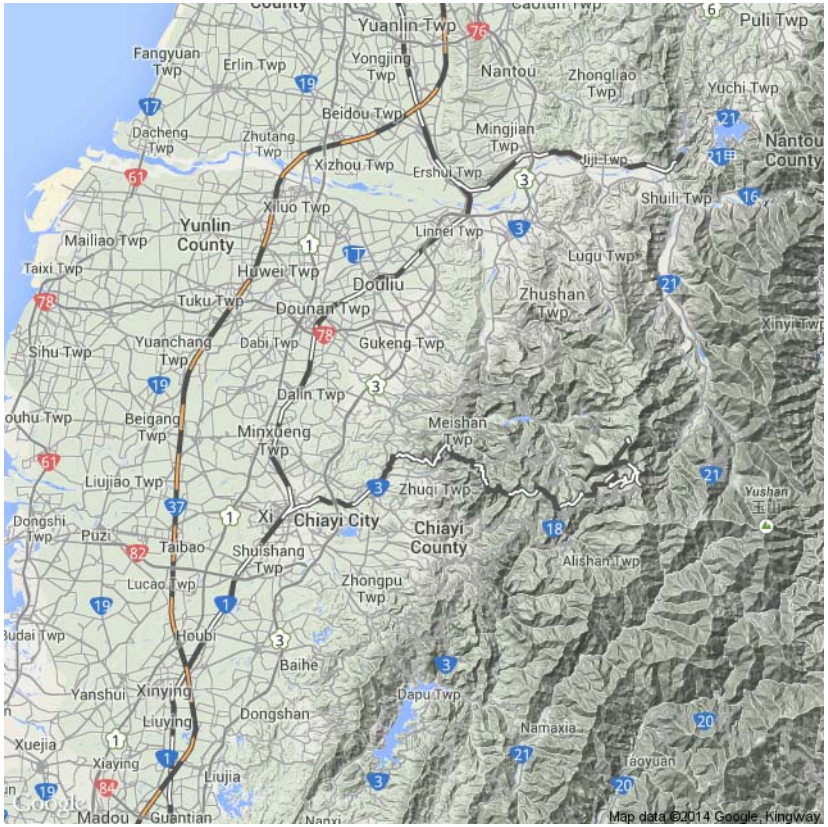
Plotting on Google Static Maps in R: RgoogleMaps


```
GetMap(center = c(lat = 42, lon = -76), size = c(640, 640), destfile,
    zoom = 12, markers, path = "", span, frame, hl, sensor = "true",
    maptype = c("roadmap", "mobile", "satellite", "terrain",
        "hybrid", "mapmaker-roadmap", "mapmaker-hybrid")[2],
    format = c("gif", "jpg", "jpg-baseline", "png8", "png32")[5],
    RETURNIMAGE = TRUE, GRAYSCALE = FALSE, NEWMAP = TRUE, SCALE = 1,
    verbose = 0)
```


```
library(RgoogleMaps)
WorldMap <- GetMap(center=c(0,0), zoom =1,
                   destfile = "World1.png")
```

台灣地圖

```
TaiwanMap <- GetMap(center=c(lat = 23.58, lon =120.58), zoom =7, destfile =
"Taiwan1.png")
TaiwanMap <- GetMap(center=c(lat = 23.58, lon =120.58), zoom = 10, destfile =
"Taiwan2.png", maptype = "terrain")
```

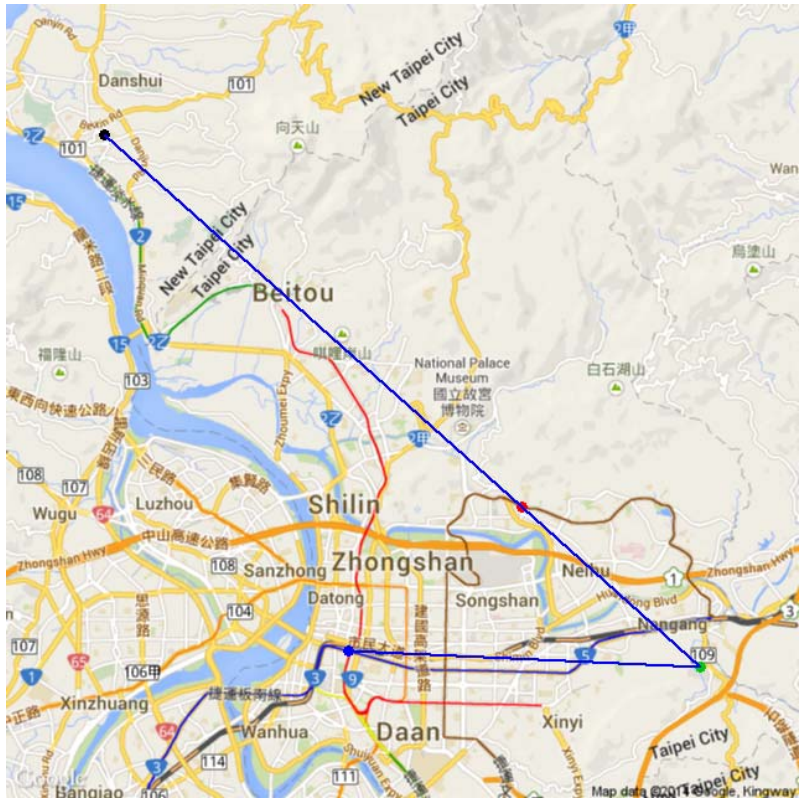


於地圖上標記

```
my.lat <- c(25.175339, 25.082288, 25.042185, 25.046254)
my.lon <- c(121.450003, 121.565481, 121.614548, 121.517532)
bb = qbbox(my.lat, my.lon)
print(bb)
MyMap <- GetMap.bbox(bb$lonR, bb$latR, destfile = "my.png", maptype = "roadmap")

My.markers <- cbind.data.frame(lat = my.lat, lon = my.lon)
tmp <-  PlotOnStaticMap(MyMap, lat = My.markers[,"lat"], lon = My.markers[,"lon"], destfile =
"my.png", cex=2.5, pch=20, col=1:4, add=F)
```

查詢經緯度
http://card.url.com.tw/realads/map_latlng.php

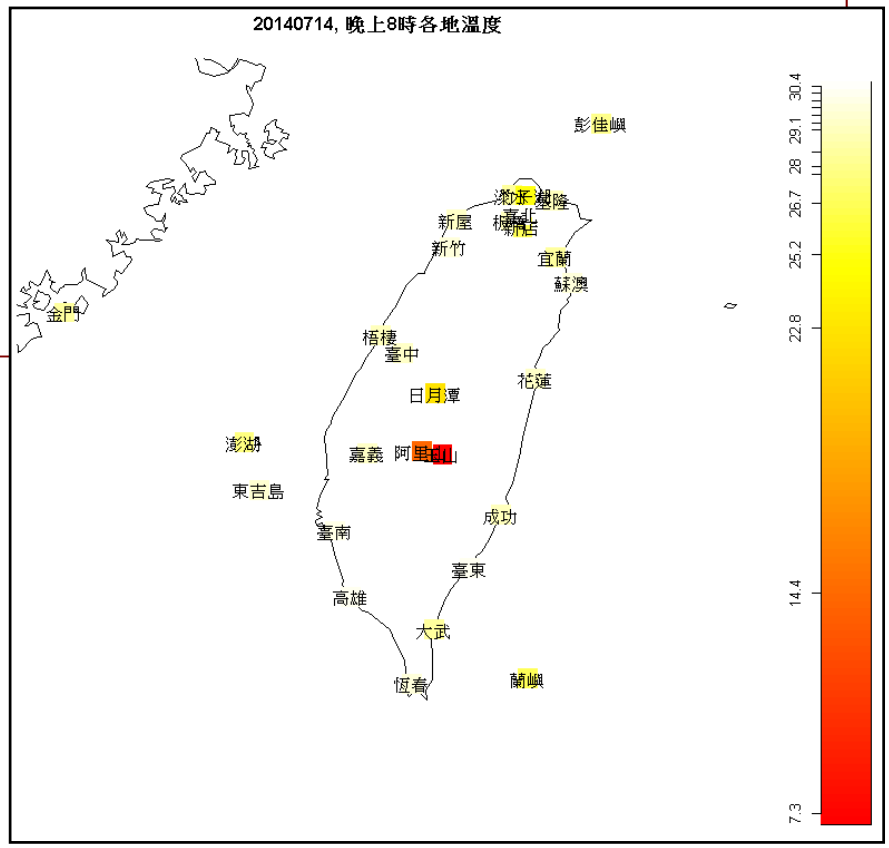
- 淡江大學 25.175339, 121.450003
- 台北市的地理中心位置: 內湖區環山路和內湖路一段跟基湖路口: 25.082288, 121.565481
- 中研院 25.042185, 121.614548
- 捷運台北站: 25.046254,121.517532



Package: maps, mapdata

氣象局開放資料平台 http://opendata.cwb.gov.tw/

```
library(maps); library(maptools); library(mapdata); library(mapproj)
layout(matrix(c(1,1,1,0,2,0), ncol=2), widths=c(10, 1), heights=c(1, 10, 1))
map("world2Hires", xlim=c(118, 123), ylim=c(21, 26))
data <- read.table("20140714-weather.txt", sep="\t", header=TRUE, row.names=1)
x <- data$TEMP
tm <- floor((100-1)/(max(x)-min(x))*(x-min(x)) + 1)
used.col <- heat.colors(100)[tm]
points(data$lon, data$lat, pch=15, col=used.col)
text(data$lon, data$lat, labels=row.names(data))
title("20140714, 晚上8時各地溫度")
par(mar=c(1,1,1,1))
image(t(matrix(c(1:100), ncol=1)),
      col=heat.colors(100), xaxt="n", yaxt="n")
axis(LEFT <- 2, at=tm/100,
     labels=as.character(x), cex.axis=1)
```

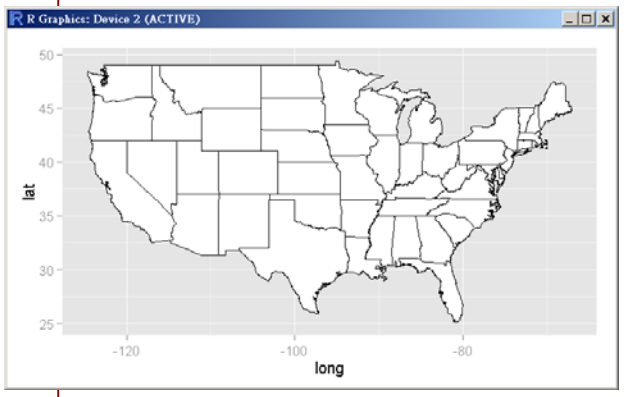


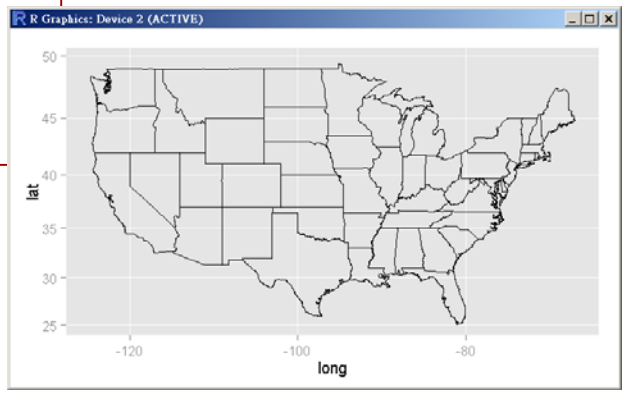
See also:

Spatial data in R: Using R as a GIS
http://pakillo.github.io/R-GIS-tutorial/

Creating a Map

```
> library(ggplot2)
> library(maps)
> library(mapproj)
> states.map <- map_data("state")
> head(states.map, 3)
       long      lat group order  region subregion
1 -87.46201 30.38968     1     1 alabama      <NA>
2 -87.48493 30.37249     1     2 alabama      <NA>
3 -87.52503 30.37249     1     3 alabama      <NA>
> tail(states.map, 3)
           long      lat group order  region subregion
15597 -107.9223 41.01805    63 15597 wyoming      <NA>
15598 -109.0568 40.98940    63 15598 wyoming      <NA>
15599 -109.0511 40.99513    63 15599 wyoming      <NA>

> ggplot(states.map, aes(x=long, y=lat, group=group)) +
    geom_polygon(fill="white", colour="black")

> ggplot(states.map, aes(x=long, y=lat, group=group)) +
    geom_path() + coord_map("mercator")
```





mercator: equally spaced straight meridians, conformal, straight compass courses

Source: 13.17. Creating a Map, R Graphics Cookbook 2nd

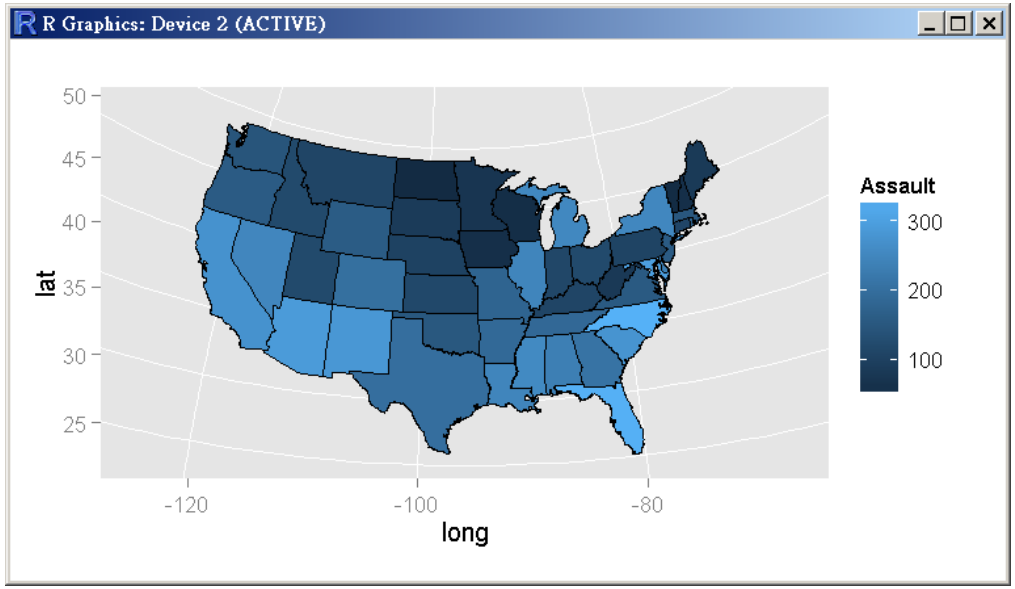
分級著色圖 (Choropleth Maps)

Violent Crime Rates by US State (USArrests): the data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

```
> head(USArrests, 3)
           Murder Assault UrbanPop Rape
Alabama      13.2     236       58 21.2
Alaska       10.0     263       48 44.5
Arizona       8.1     294       80 31.0
> crimes <- data.frame(state = tolower(rownames(USArrests)), USArrests)
> head(crimes, 3)
               state Murder Assault UrbanPop Rape
Alabama      alabama   13.2     236       58 21.2
Alaska        alaska   10.0     263       48 44.5
Arizona      arizona    8.1     294       80 31.0

> library(maps); library(ggmap)
> states.map <- map_data("state")
> head(states.map, 3)
       long      lat group order  region subregion
1 -87.46201 30.38968     1     1 alabama      <NA>
2 -87.48493 30.37249     1     2 alabama      <NA>
3 -87.52503 30.37249     1     3 alabama      <NA>
> crime.map <- merge(states.map, crimes, by.x="region", by.y="state")
> head(crime.map, 3)
   region      long      lat group order subregion Murder Assault UrbanPop Rape
1 alabama -87.46201 30.38968     1     1      <NA>   13.2     236       58 21.2
2 alabama -87.48493 30.37249     1     2      <NA>   13.2     236       58 21.2
3 alabama -87.95475 30.24644     1    13      <NA>   13.2     236       58 21.2
```

分級著色圖 (Choropleth Maps)

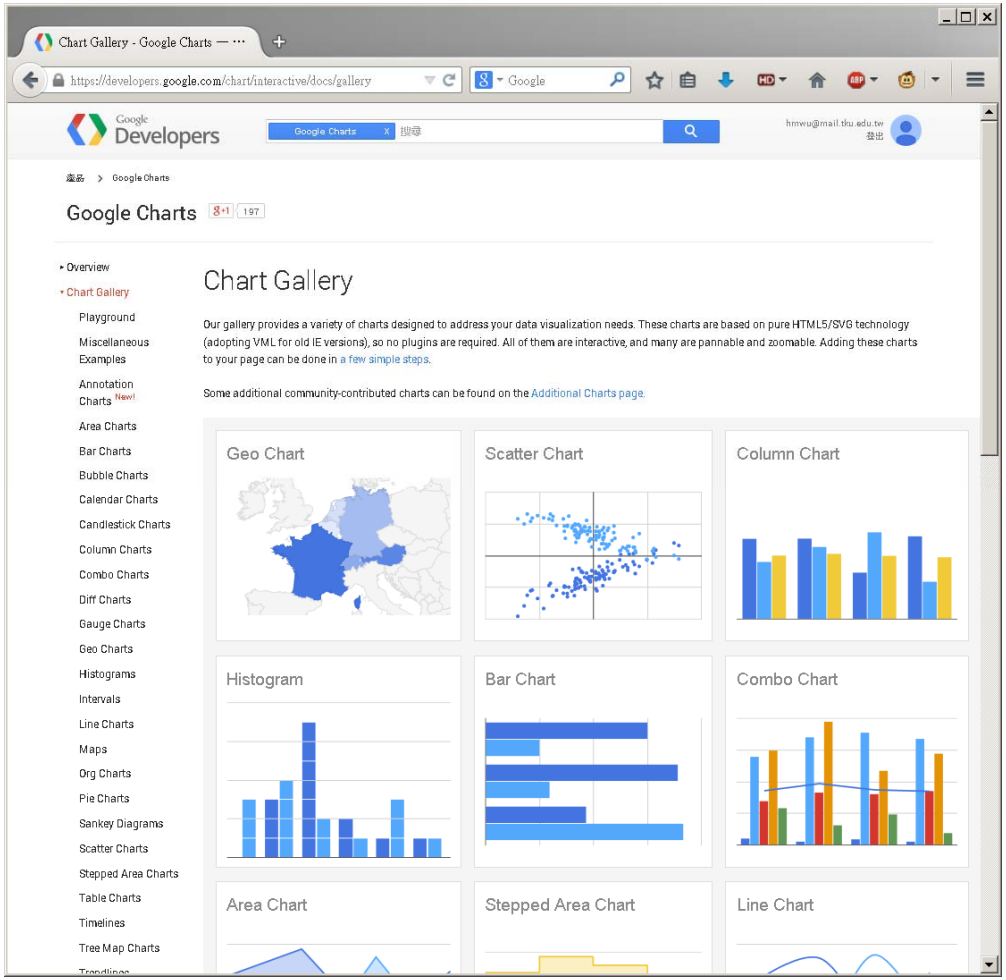
```
> # After merging, the order has changed, which would lead to polygons drawn in
> # the incorrect order. So, we sort the data.
> library(plyr) # For arrange() function
> # Sort by group, then order
> crime.map <- arrange(crime.map, group, order)
> head(crime.map, 3)
   region      long      lat group order subregion Murder Assault UrbanPop Rape
1 alabama -87.46201 30.38968     1     1      <NA>   13.2     236       58 21.2
2 alabama -87.48493 30.37249     1     2      <NA>   13.2     236       58 21.2
3 alabama -87.52503 30.37249     1     3      <NA>   13.2     236       58 21.2
> ggplot(crime.map, aes(x=long, y=lat, group=group, fill=Assault)) +
        geom_polygon(colour="black") +
        coord_map("polyconic")
```

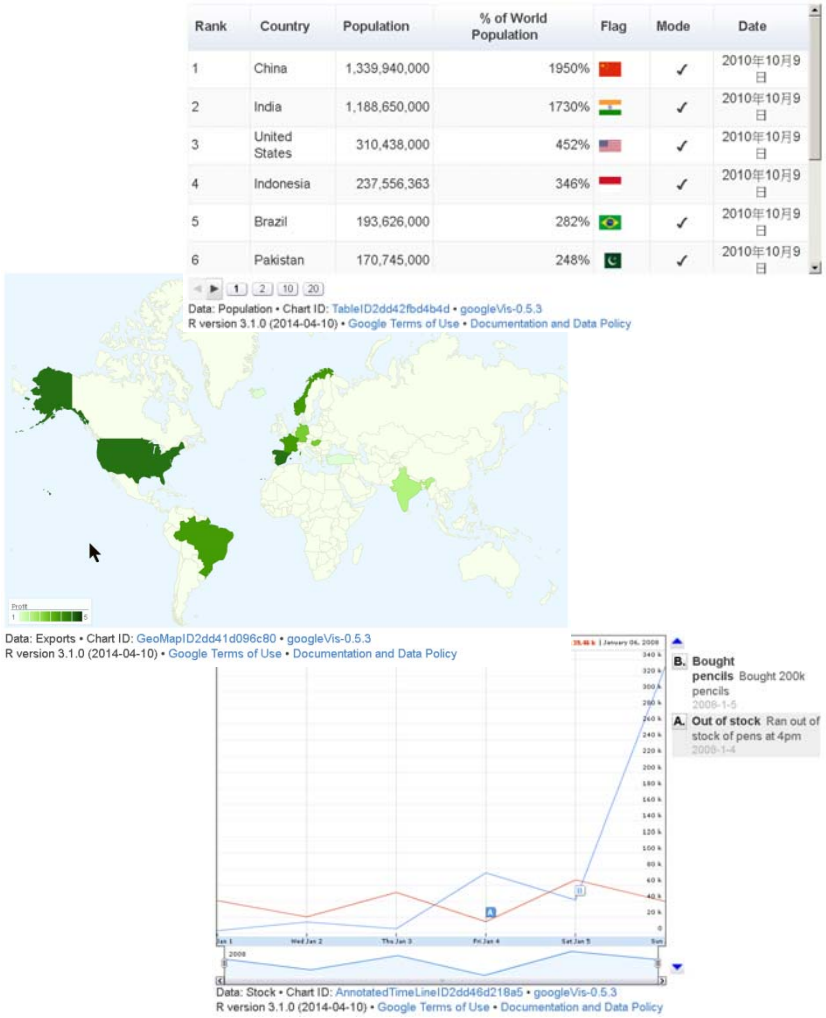


Package: googleVis

Interface between R and Google Charts

```
library(googleVis)
demo(googleVis)
```

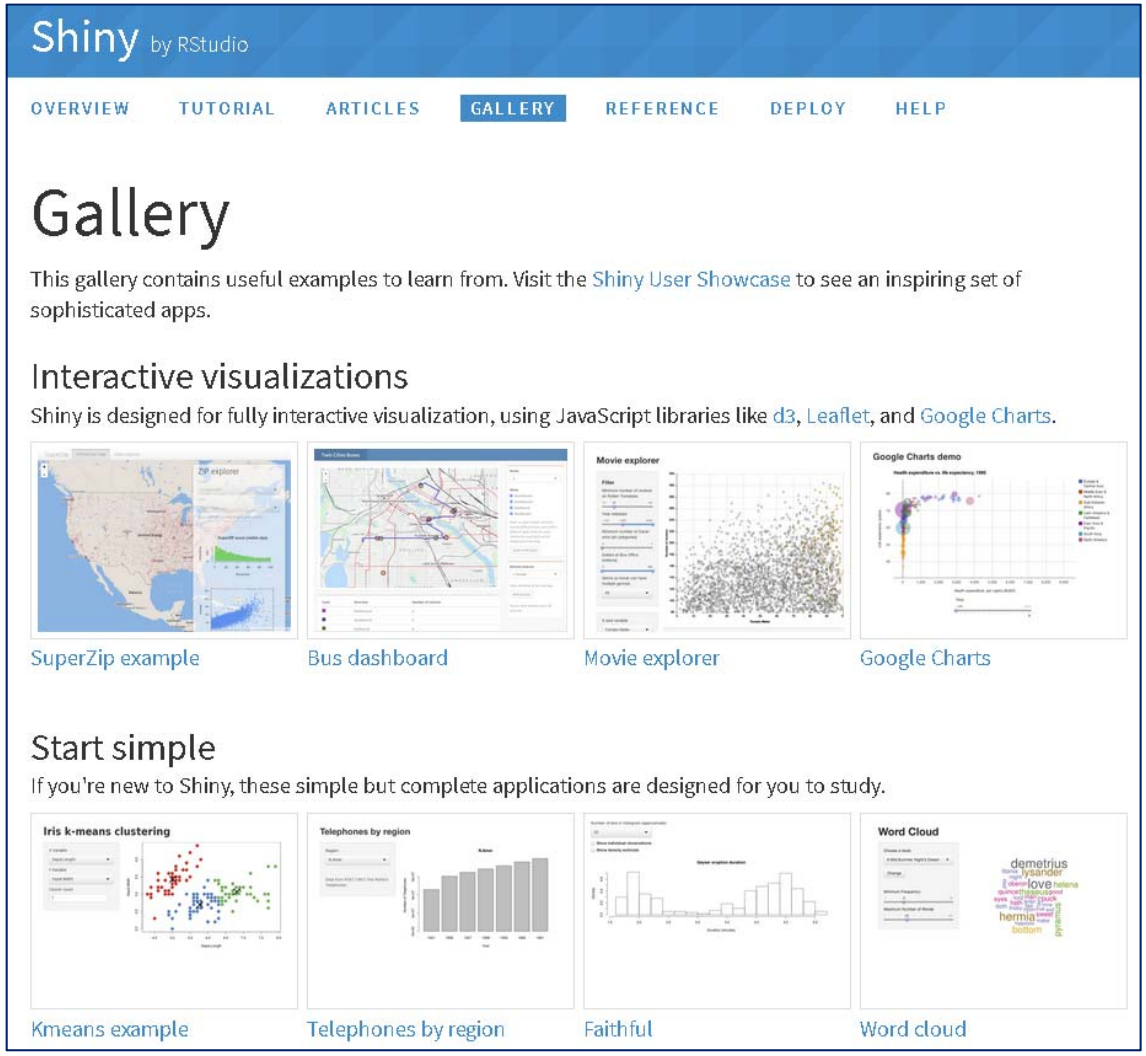




http://cran.r-project.org/web/packages/googleVis/vignettes/googleVis_examples.html

Shiny

A web application framework for R



http://shiny.rstudio.com/

Turn analyses into interactive web applications

- Widgets: Sliders, File Download
- Application layout
- Dynamic user interface
- Interactive plots
- ...

ggvis: Interactive Grammar of Graphics

ggvis 0.4 overview

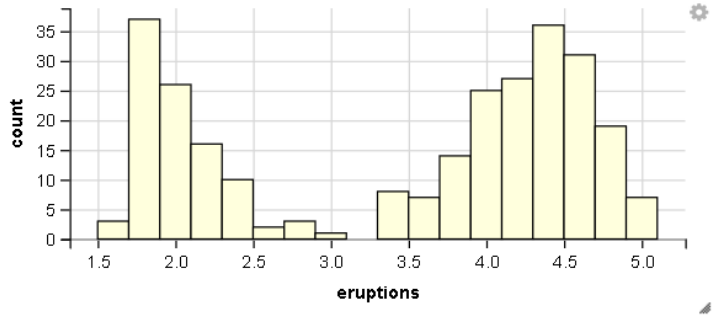
ggvis is a data visualization package for R which lets you:

- Declaratively describe data graphics with a syntax similar in spirit to ggplot2.
- Create rich interactive graphics that you can play with locally in Rstudio or in your browser.
- Leverage shiny's infrastructure to publish interactive graphics usable from any browser (eith

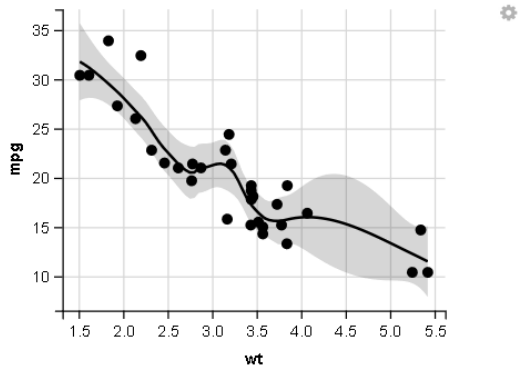
The goal is to combine the best of R (e.g. every modelling function you can imagine) and the be manipulation and transformation are done in R, and the graphics are rendered in a web browser display in a viewer panel, which is possible because RStudio is a web browser.

Examples of ggvis graphics

Histogram:



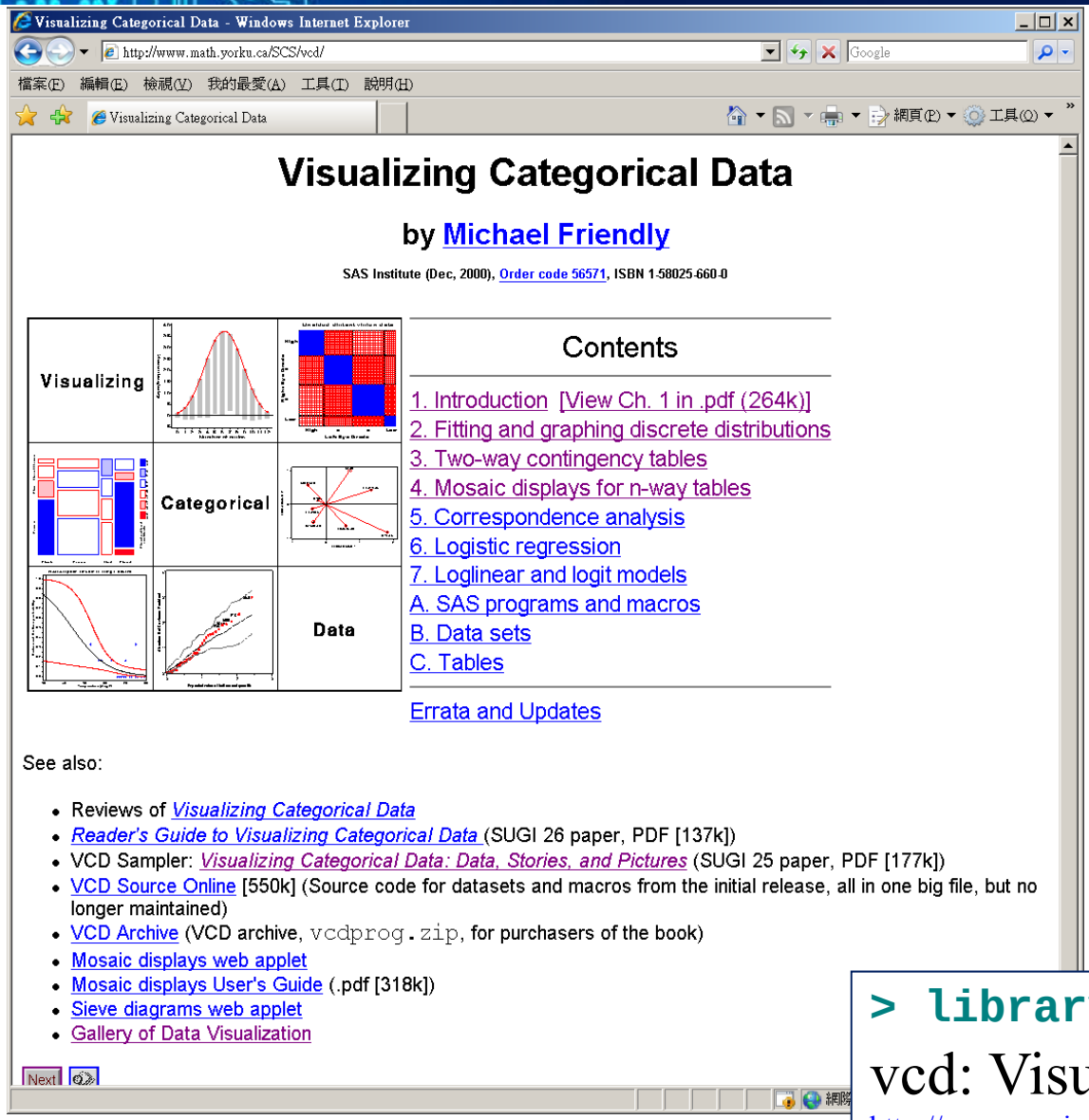
Scatterplot with smooth curve and interactive control:



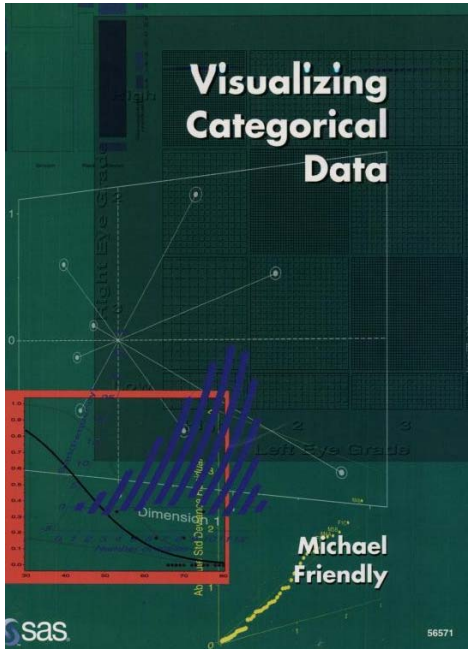
Smoothing span

An implementation of an interactive grammar of graphics, taking the best parts of 'ggplot2', combining them with the reactive framework of 'shiny' and drawing web graphics using 'vega'.

類別資料的視覺化: vcd
Visualizing Categorical Data



- Fourfold Display for 2x2 Tables
- Association Plots
- Mosaic Display



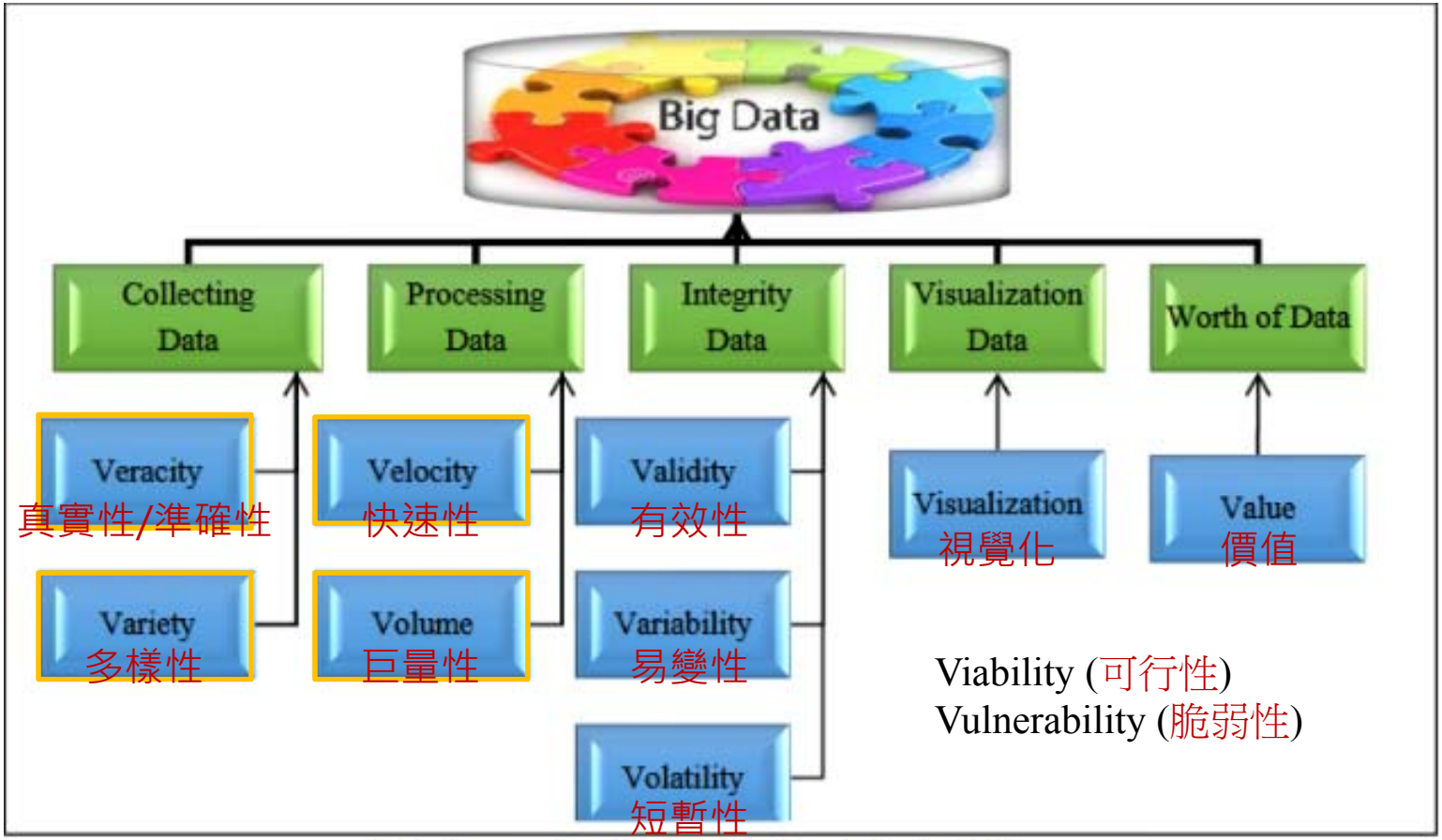

```
> library(vcd)
```

vcd: Visualizing Categorical Data

http://cran.r-project.org/web/packages/vcd/index.html

Big Data: The Era of 9 Vs

- Visualization:
  - Visualization will be key to making big data an integral part of decision making.
  - Visualization will be the only way to make big data accessible to a large audience.
  - Visualization will be essential to the analysis of big data so it can be of highest value.



Categorization of Big Data V's

http://blogs.systweak.com/2017/03/big-data-vs-represents-characteristics-or-challenges-of-big-data/

Big Data Visualization

- **Definition - What does Big Data Visualization mean?**
  - Big data visualization refers to the implementation of more contemporary visualization techniques to illustrate the relationships within data. Visualization tactics include applications that can display real-time changes and more illustrative graphics, thus going beyond pie, bar and other charts. These illustrations veer away from the use of hundreds of rows, columns and attributes toward a more artistic visual representation of the data.
- **Techopedia explains Big Data Visualization**
  - Normally when businesses need to present relationships among data, they use graphs, bars and charts to do it. They can also make use of a variety of colors, terms and symbols. The main problem with this setup, however, is that it doesn't do a good job of presenting very large data or data that includes huge numbers. Data visualization uses more interactive, graphical illustrations - including personalization and animation - to display figures and establish connections among pieces of information.

techopedia™

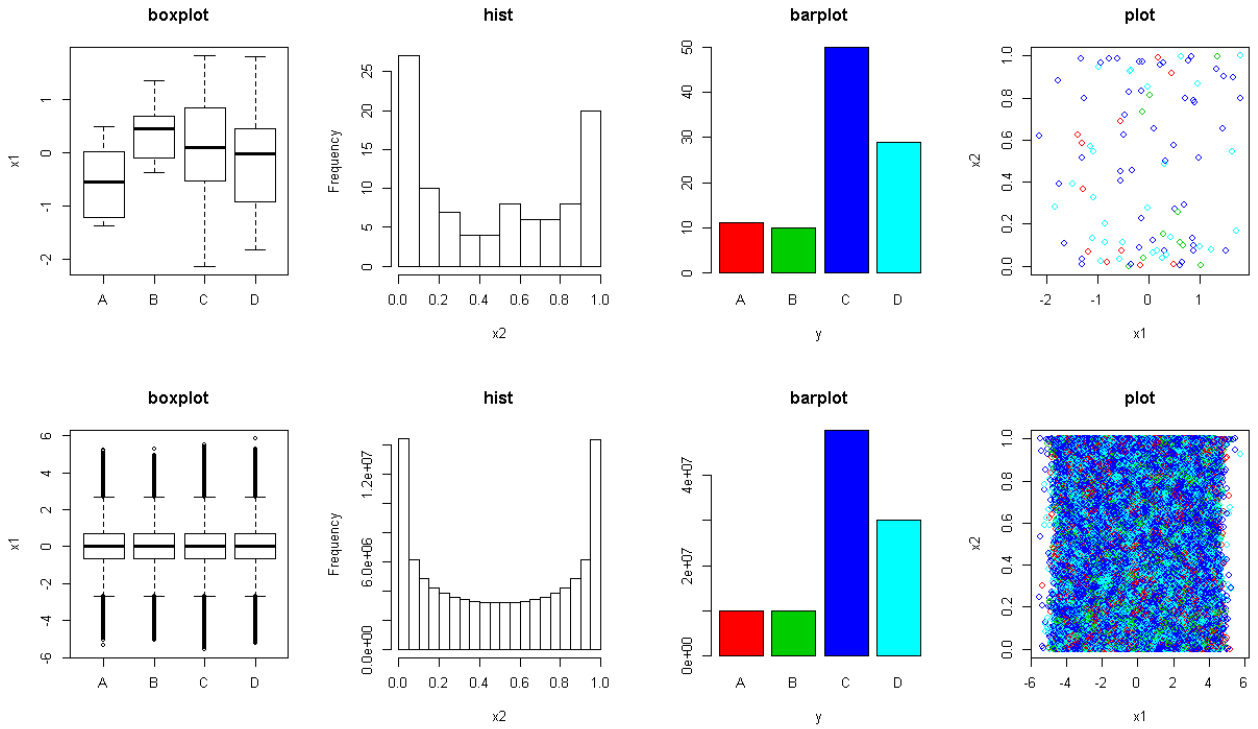
https://www.techopedia.com/definition/28988/big-data-visualization

The Challenge of Visualizing Big Data

```
> n <- 1e+02
```

a large p?

```
> n <- 1e+08
```




```
> n <- 1e+02
> y <- as.factor(sample(LETTERS[1:4], n, replace=T, prob=c(0.1, 0.1, 0.5, 0.3)))
> x1  <- rnorm(n)
> x2  <- rbeta(n, 0.5, 0.5)
> xydata <- data.frame(y, x1, x2)
> par(mfrow=c(1,4))
> boxplot(x1~y, data=xydata, ylab="x1", main="boxplot")
> hist(x2, xlab="x2", main="hist")
> barplot(table(y), xlab="y", col = 2:5, main="barplot")
> plot(x1, x2, main="plot", col=as.integer(y)+1)
```

Two principles:
Look at Less Data;
or Look at Data Faster

Graphics of Large Datasets

Scatterplot Matrix Techniques for Large *N*

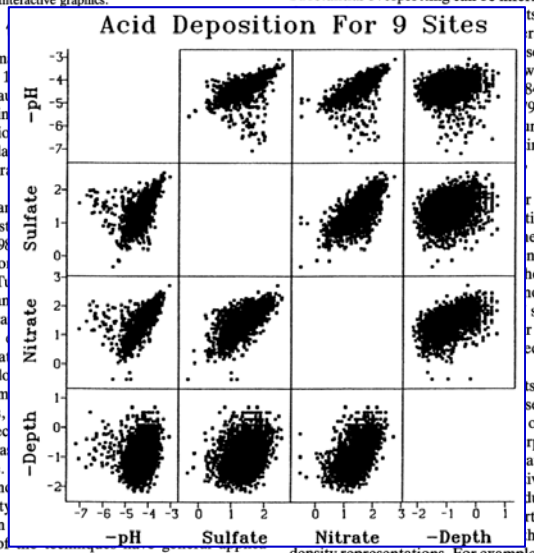
D. B. CARR, R. J. LITTLEFIELD, W. L. NICHOLSON, and J. S. LITTLEFIELD*

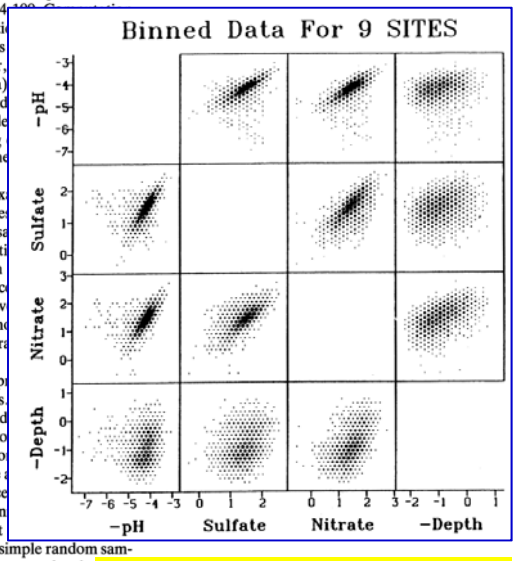
High-performance interaction with scatterplot matrices is a powerful approach to exploratory multivariate data analysis. For a small number of data points, real-time interaction is possible and overplotting is usually not a major problem. When the number of plotted ... ever, display techniques that deal with overplotting ... are important. This article addresses these two prob... density representation by gray scale or by symbol ... brushing, and animation sequences. We also discus... generally applicable, including interactive graphical ... any plot in a collection of scatterplots and comp... matrices.

KEY WORDS: Density representations; Animation sequences; Graphical subset selection; Interactive graphics.

500 points can seem large. For our purposes, *N* is large if plotting or computation times are long, or if plots can have an extensive amount of overplotting. Figure 1 provides an ... large. The exhibit also fits the other definitions of large. Substantial overplotting can be inferred from plotted area, ... density representations. For example, simple random sampling omits most outliers and obscures some local ... while introducing others. In other words, sampling ... alter features that are frequently of interest in ter... subsetting, identification, and comparison. Sampling ... powerful tool, but even sophisticated analysts wh... handle complex sampling will find it useful to look ... whole data set.

each scatterplot would contain 4,109 points. With 14 plots, the total number of points in the display is almost 50,000.





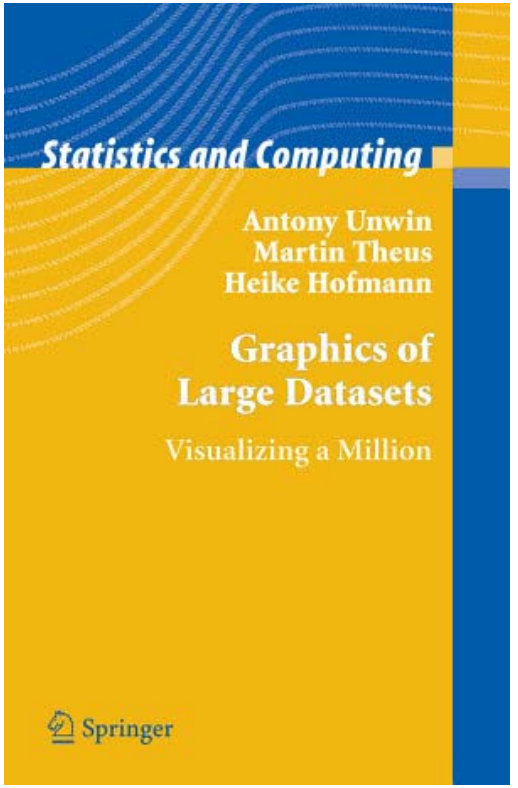
Hexagon area density representation

2. LARGE *N*

What is large depends on the frame of reference. If available plotting space for a scatterplot is a one-inch square, ...

* D. B. Carr is Senior Research Scientist, Statistics; R. J. Littlefield is Senior Research Scientist, Computers and Information Sciences; W. L. Nicholson is Lead Scientist, Statistics; and J. S. Littlefield is Research Scientist, Statistics; all in the Computational Sciences Department, Pacific Northwest Laboratory, Richland, WA 99352. The authors would like to thank Vern Crow for substantial software contributions and Paul Tukey for suggesting the hexagonal binning algorithm in the Appendix. This work was supported by the Applied Mathematical Sciences division of the U.S. Department of Energy under Contract DE-ACO6-76RLO 1830.

Journal of the American Statistical Association
June 1987, Vol. 82, No. 398, Statistical Graphics



Antony Unwin, Martin Theus, Heike Hofmann, Publisher: Springer; 2006 edition (July 24, 2006)

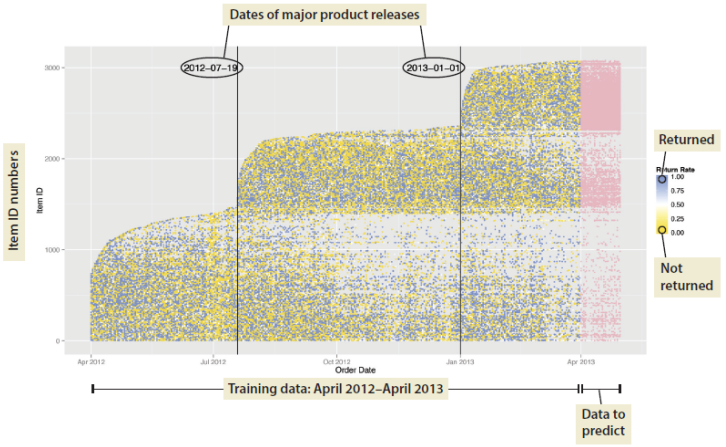
Data Visualization and Statistical Graphics in Big Data Analysis

ANNUAL REVIEWS
For Librarians & Agents For Authors
JOURNALS A-Z MULTIMEDIA JOURNAL INFO PRICING & SU
Home / Annual Review of Statistics and Its Application / Volume 3, 2016 / Cook, pp 133-159

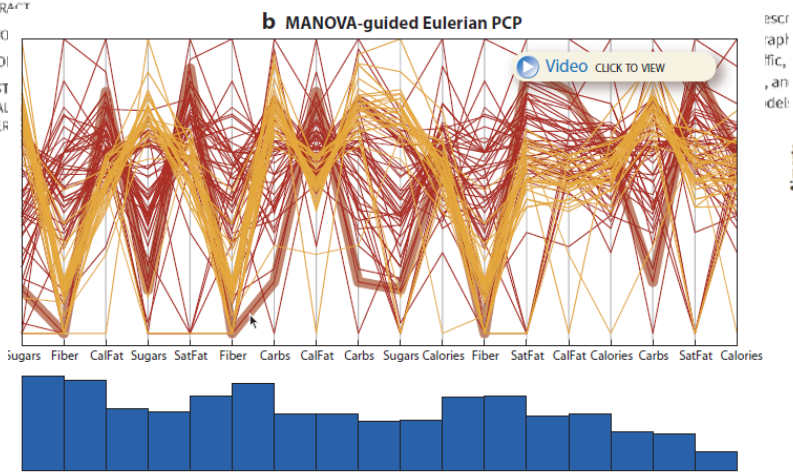
Data Visualization and Statistical Graphics in Big Data Analysis

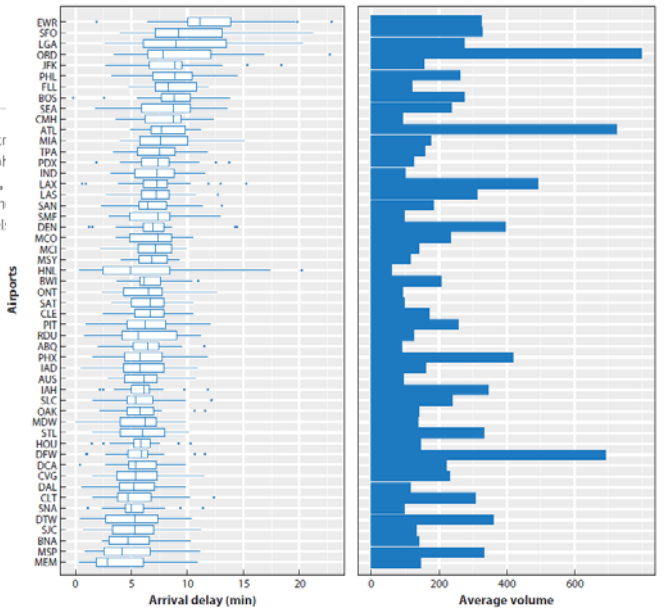
Annual Review of Statistics and Its Application
Vol. 3:133-159 (Volume publication date June 2016)
DOI: 10.1146/annurev-statistics-041715-033420

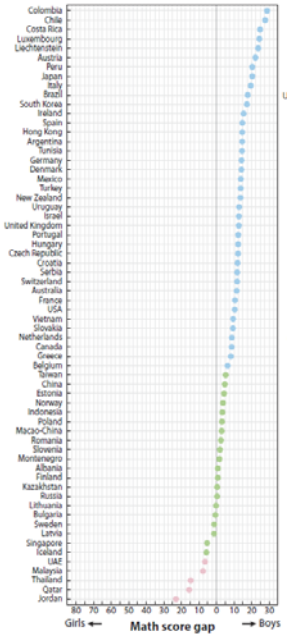
Dianne Cook,[1] Eun-Kyung Lee,[2] and Mahbubul Majumder[3]
[1]Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia; email: dicook@monash.edu
[2]Department of Statistics, Ewha Womans University, Seoul 120-750, Korea; email: lee.eunk@ewha.ac.kr
[3]Department of Mathematics, University of Nebraska, Omaha, Nebraska 68182; email: mmajumder@unomaha.edu



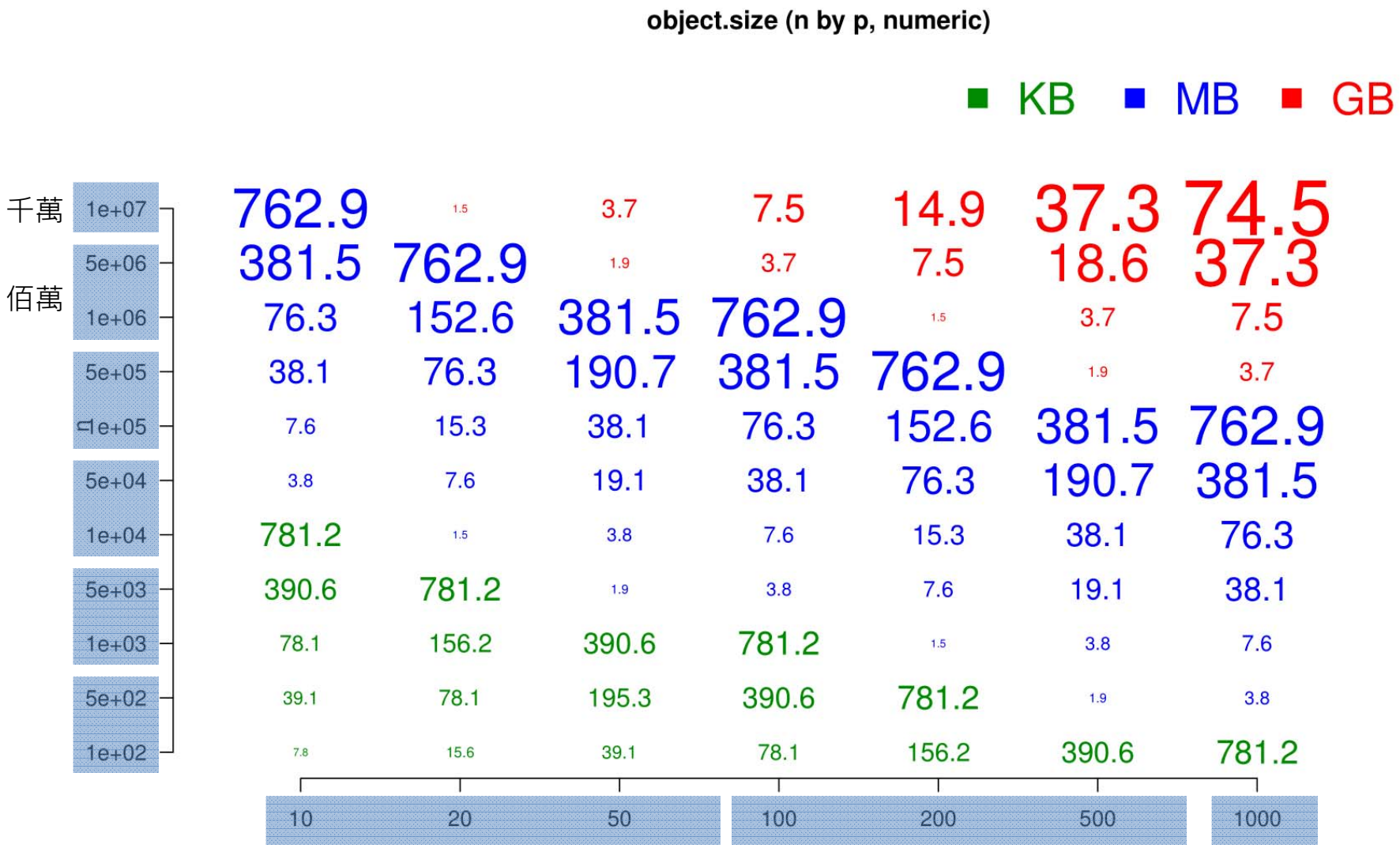
Abstract







object.size{utils}

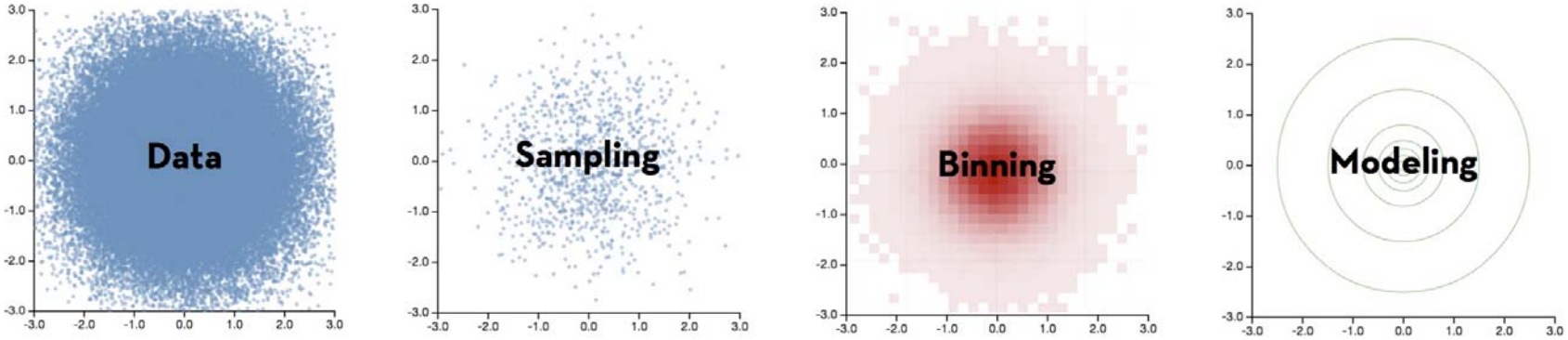


(n*p*8)/(1024*1024) MB

1 Bit = Binary Digit; 8 Bits = 1 Byte; 1024 Bytes = 1 Kilobyte; 1024 Kilobytes = 1 Megabyte
1024 Megabytes = 1 Gigabyte; 1024 Gigabytes = 1 Terabyte; 1024 Terabytes = 1 Petabyte

How Can We Visualize and Interact with Billion+ Record Databases in Real-time?

- Two central challenges:
  - need to keep visualizations perceptually effective regardless of the number of input data points.
  - need to support real-time interaction to enable rapid and iterative exploratory analysis.
- Perceptual and interactive scalability should be limited by the chosen resolution of the visualized data, not the number of records.



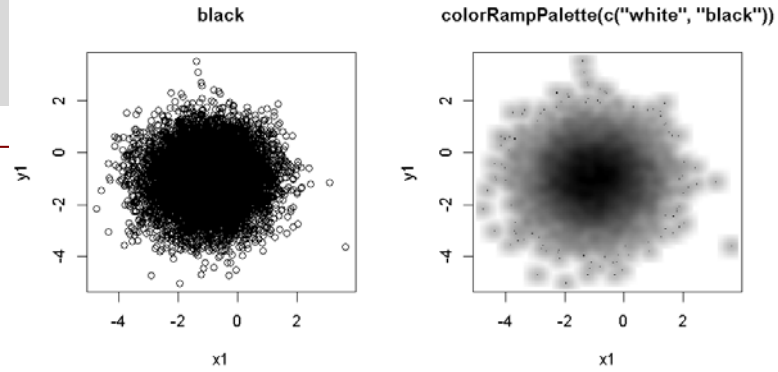
http://skandel.github.io/slides/strata2013/part1

smoothScatter {graphics}: Scatterplots with Smoothed Densities Color Representation

- **smoothScatter** produces a smoothed color density representation of a scatterplot, obtained through a (2D) kernel density estimate.
  - **nbin**: numeric vector of length one (for both directions) or two (for x and y separately) specifying the number of equally spaced grid points for the density estimation; directly used as gridsize in bkde2D().
  - **bandwidth**: numeric vector (length 1 or 2) of smoothing bandwidth(s). If missing, a more or less useful default is used. bandwidth is subsequently passed to function bkde2D.
  - **nrpoints**: number of points to be superimposed on the density image. The first nrpoints points from those areas of lowest regional densities will be plotted. Adding points to the plot allows for the identification of outliers. If all points are to be plotted, choose nrpoints = Inf.

```
smoothScatter(x, y = NULL, nbin = 128, bandwidth,
              colramp = colorRampPalette(c("white", blues9)),
              nrpoints = 100, ret.selection = FALSE,
              pch = ".", cex = 1, col = "black",
              transformation = function(x) x^.25,
              postPlotHook = box,
              xlab = NULL, ylab = NULL, xlim, ylim,
              xaxs = par("xaxs"), yaxs = par("yaxs"),
```

```
> n <- 1e+04
> x1 <- rnorm(n, mean = -1, sd = 1)
> y1 <- rnorm(n, mean = -1, sd = 1)
> x2 <- rnorm(n, mean = 2, sd = 1)
> y2 <- rnorm(n, mean = 2, sd = 1)
> par(mfrow=c(1, 2))
> plot(x1, y1, main="black")
> smoothScatter(x1, y1, col="black", colramp=colorRampPalette(c("white", "black")),
main='colorRampPalette(c("white", "black"))')
```

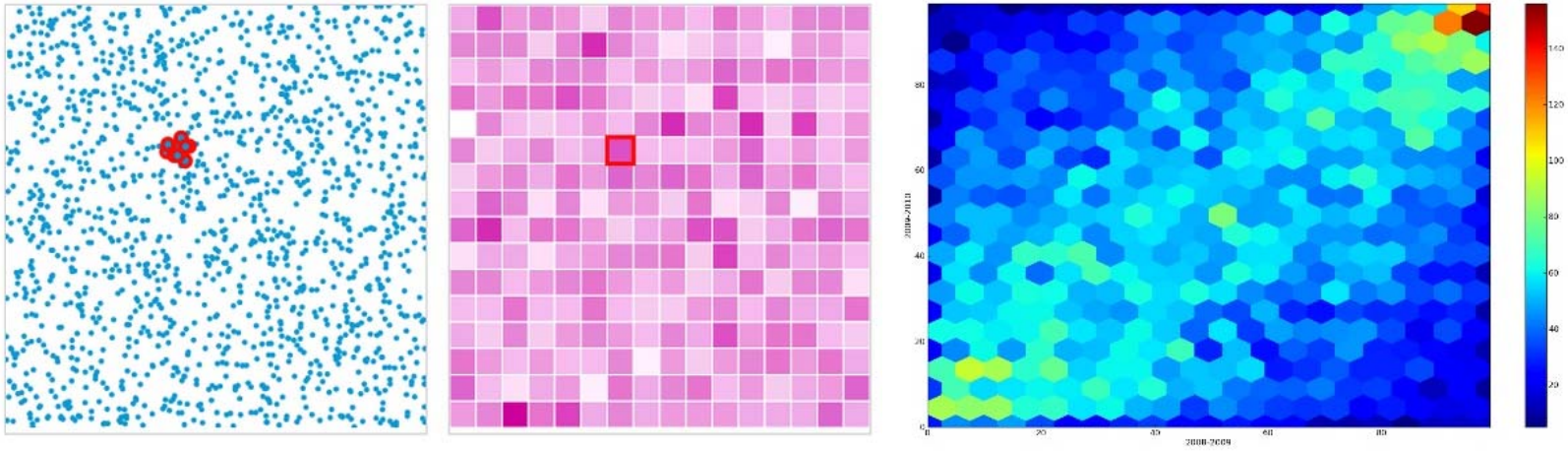


Binning Technique

- Binning is a technique of data aggregation used for grouping a dataset of N values into less than N discrete groups.
  - the XY plane is uniformly tiled with polygons (squares, rectangles or hexagons).
  - the number of points falling in each bin (tile) are counted and stored in a data structure.
  - the bins with count > 0 are plotted using a color range (heatmap) or varying their size in proportion to the count.

Rectangular binning

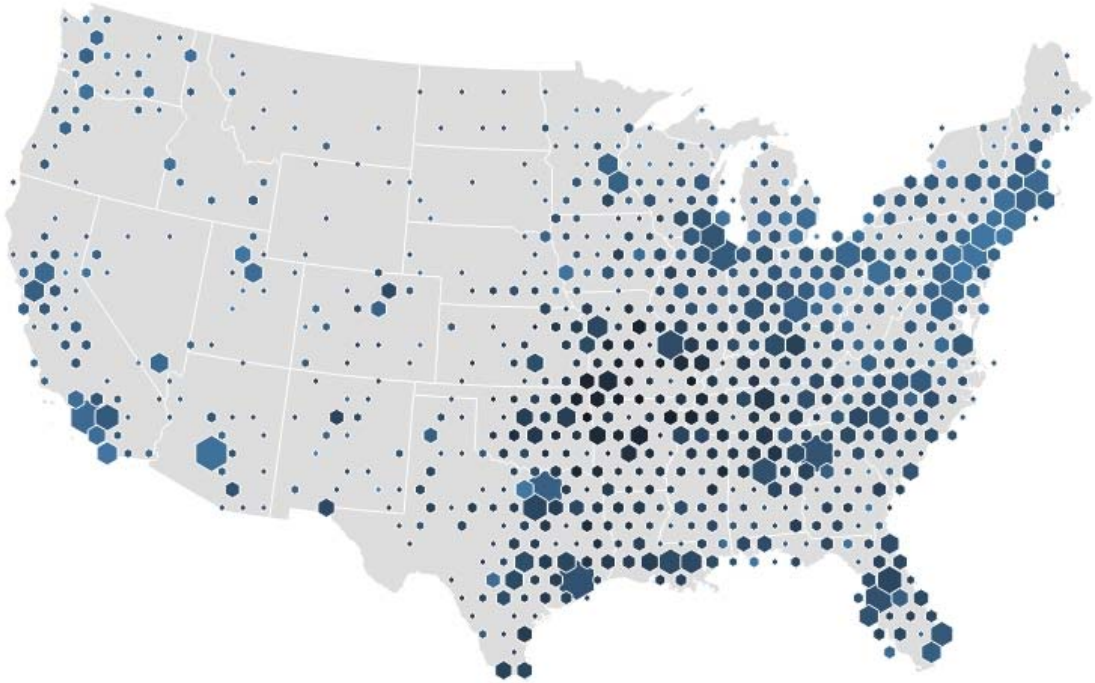
Hexagonal binning



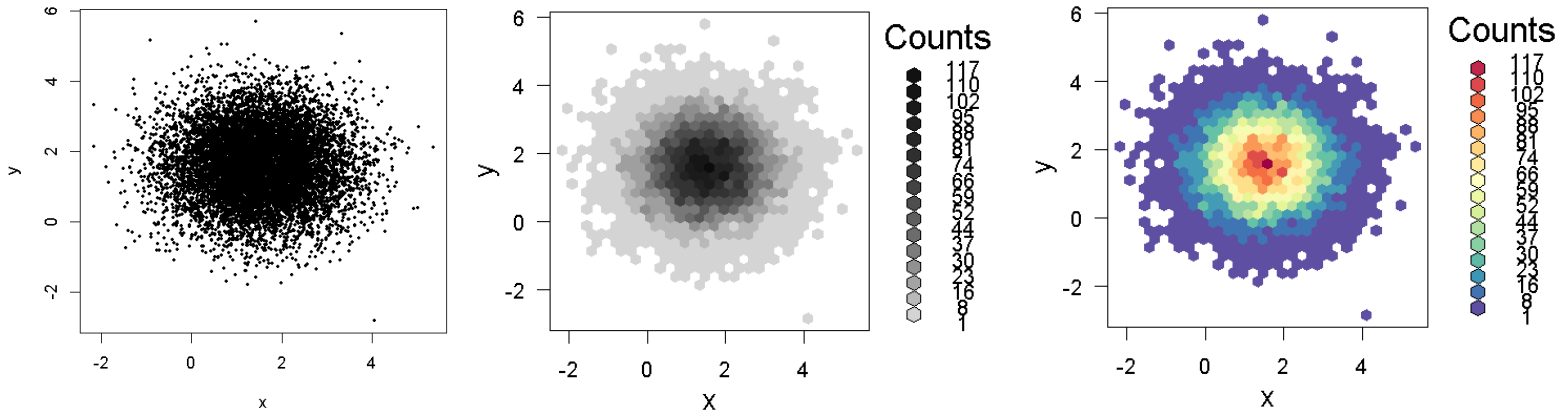
http://www.meccanismocomplesso.org/en/hexagonal-binning/

Why hexagons?

- Working over a larger area, a square grid will suffer more from distortion due to curvature than hexagons.
- Hexagons have symmetry of nearest neighbors which is lacking in square bins.
- Hexagons are visually less biased for displaying densities than other regular tesselations.
- The hexagon is the most complex regular polygon that can fill a plane (without gaps or overlap).

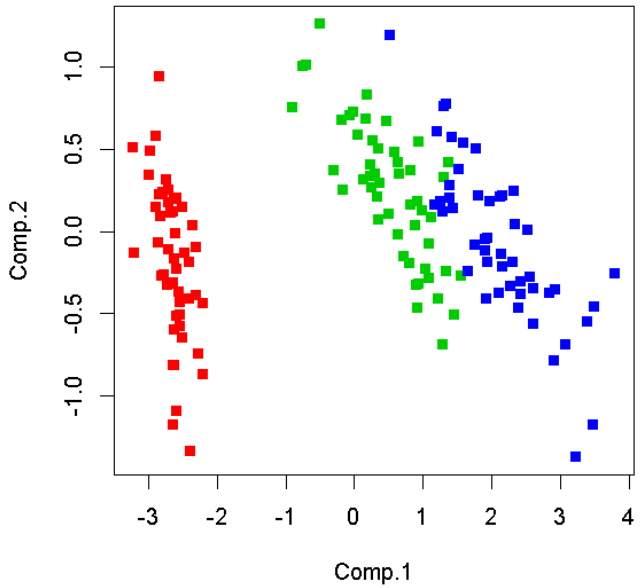


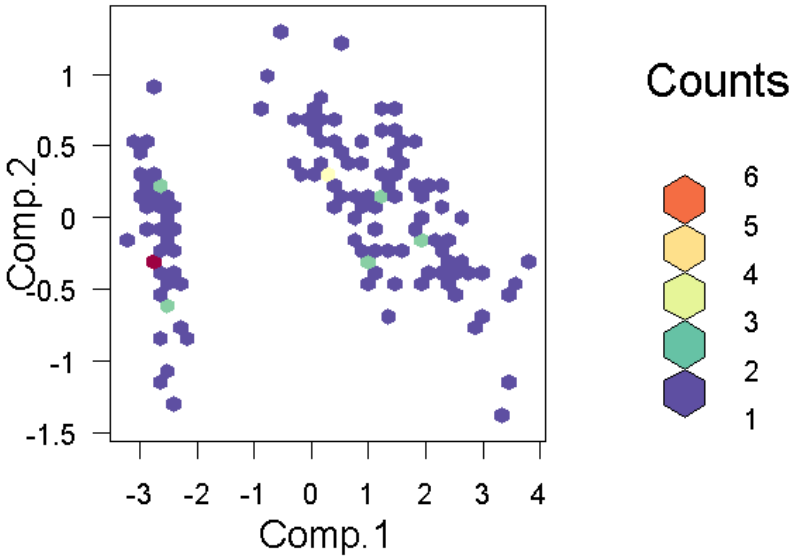
hexbin Package: Hexagonal Binning Routines




```
> x <- rnorm(mean=1.5, 10000)
> y <- rnorm(mean=1.6, 10000)
> my.data <- data.frame(x, y)
>
> pk <- c("RColorBrewer", "hexbin", "gplots")
> install.packages(pk, repos="http://cran.csie.ntu.edu.tw")
> library(RColorBrewer)
> # create rainbow color
> col_rb <- colorRampPalette(rev(brewer.pal(11, 'Spectral')))
> # scatterplot
> plot(my.data, pch=16, col='black', cex=0.5)
> library(hexbin)
> h <- hexbin(my.data) # create a hexbin object
> h
'hexbin' object from call: hexbin(x = my.data)
n = 10000  points in    nc = 598  hexagon cells in grid dimensions  36 by 31
> plot(h) # in grey level
> plot(h, colramp=col_rb) # rainbow color
```

hexbin for Small Dataset

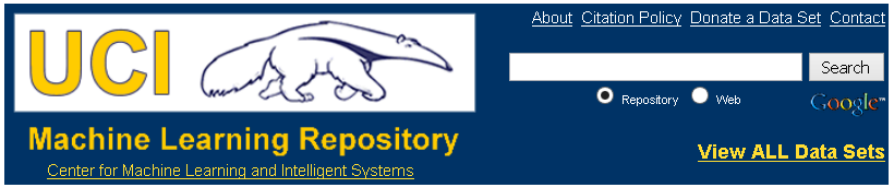





```
> iris.pca <- princomp(iris[,-5])
> plot(iris.pca$scores[,1:2], pch=15, col=as.integer(iris[,5])+1)
> iris.hex <- hexbin(iris.pca$scores[,1:2])
> plot(iris.hex, colramp=col_rb)
```

hexbin for Large Dataset

https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption



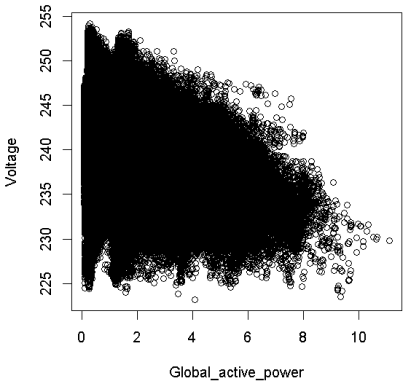
Individual household electric power consumption Data Set

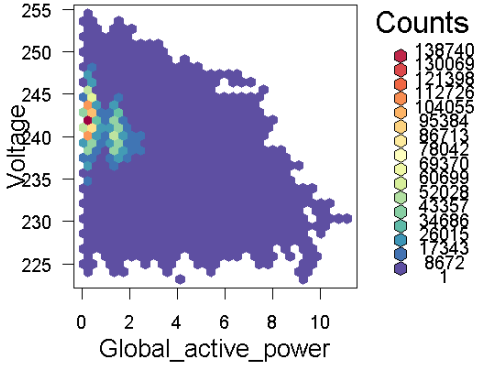
Download: Data Folder, Data Set Description

Abstract: Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available.

Data Set Characteristics:	Multivariate, Time-Series	**Number of Instances:**	2075259	**Area:**	Physical
Attribute Characteristics:	Real	**Number of Attributes:**	9	**Date Donated**	2012-08-30
Associated Tasks:	Regression, Clustering	**Missing Values?**	Yes	**Number of Web Hits:**	128715

December 2006 and November 2010 (47 months).






```
> zz <- unz(description="household_power_consumption.zip",
filename="household_power_consumption.txt")
> colC <- c(rep("NULL", 2), "numeric", "NULL", "numeric", rep("NULL", 4))
> power <- read.table(zz, header=T, sep=";", colClasses = colC, na.strings = "?")
> summary(power)
 Global_active_power    Voltage
 Min.   : 0.076      Min.   :223.2
 1st Qu.: 0.308      1st Qu.:239.0
 Median : 0.602      Median :241.0
 Mean   : 1.092      Mean   :240.8
 3rd Qu.: 1.528      3rd Qu.:242.9
 Max.   :11.122      Max.   :254.2
 NA's   :25979       NA's   :25979
```

```
> plot(power)
> power.hex <- hexbin(power)
> plot(power.hex, colramp=col_rb)
```

bigvis: Exploratory data analysis for large datasets (10-100 million observations)

- Hadley Wickham, 2013, Bin-summarise-smooth: a framework for visualising large data. https://github.com/hadley/bigvis

- The aim is to have most operations take less than 5 seconds on commodity hardware, even for 100,000,000 data points.

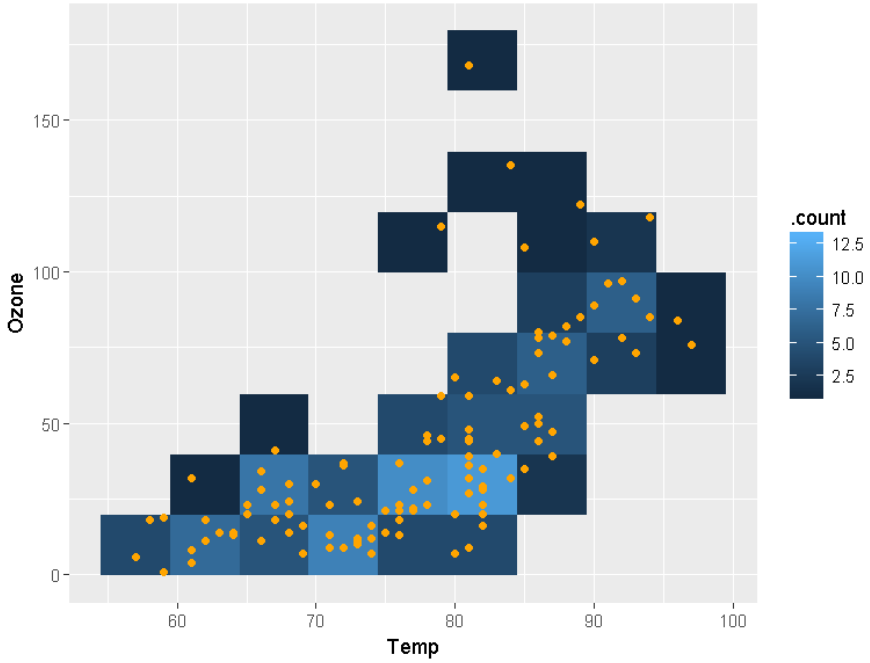
- **Workflow**:
  - **Binning**: binning is an injective mapping from the real numbers to a fixed and finite set of integers. (fixed width binning: fast, easily extended from 1d to nd).
  - **Summarizing**: to collapse the points in each bin into a small number of summary statistics. (count, sum, mean, median or sd)
  - **Smoothing**: if the estimates are rough, you might want to `smooth()`.
  - **Visualizing**: visualize the results with `autoplot {ggplot2}`

- `bigvis` provides outlier removal and smoothing:
  - big data means very rare cases can occur ⇒ outliers may be more of a problem
  - smoothing very important to highlight trends & suppress noise

```
# install.packages("devtools")
devtools::install_github("hadley/bigvis")
```

bigvis applied to a small dataset

```
> library(bigvis)
> library(ggplot2)
> head(airquality)
  Ozone Solar.R Wind Temp Month Day
1    41     190  7.4   67     5   1
2    36     118  8.0   72     5   2
3    12     149 12.6   74     5   3
4    18     313 11.5   62     5   4
5    NA      NA 14.3   56     5   5
6    28      NA 14.9   66     5   6
> par(mfrow=c(1, 2))
> hist(airquality$Ozone)
> hist(airquality$Temp)
> #ggplot(data=airquality) +
> #  geom_point(mapping = aes(x = Temp, y = Ozone)
>
> binData <- with(airquality, condense(bin(Ozone, 20), bin(Temp, 5)))
Summarising with count
> binData
   Ozone Temp .count
1     NA   57      4
2     NA   67      2
...
35 129.5   87      1
36 169.5   82      1
> ggplot(data=binData, aes(Temp, Ozone, fill=.count))+
+   geom_tile() +
+   geom_point(data=airquality, aes(fill=NULL), colour="orange")
```



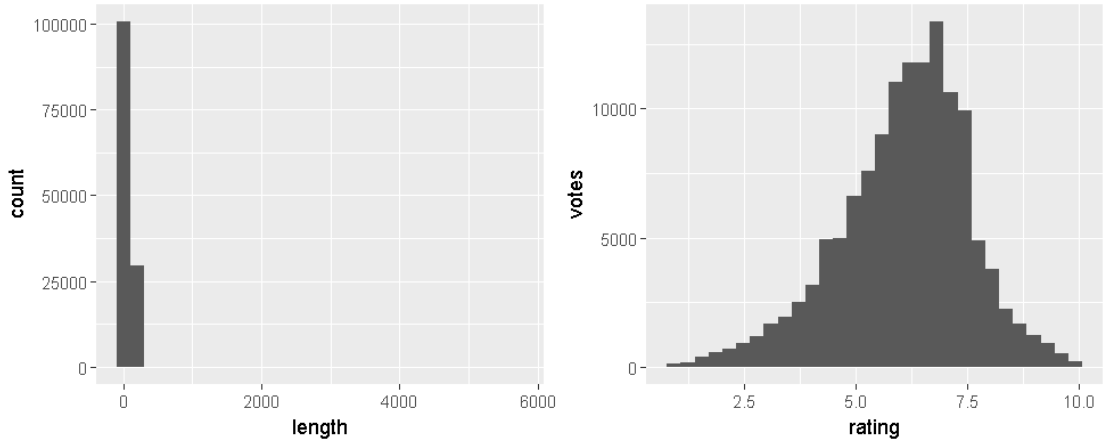
bigvis Applied to a Large Dataset with Outliers

movies {bigvis}: Movie information and user ratings from IMDB.com.

```
> data(movies)
> dim(movies)
[1] 130456     14
> head(movies)
```

```
> head(movies)
                      title year length budget rating votes mpaa Action Animation Comedy Drama Documentary Romance Short
1                Falling Cat 1890      1     NA    5.3    27 <NA>  FALSE     FALSE  FALSE FALSE       FALSE   FALSE FALSE
2          Dickson Greeting 1891      1     NA    5.8   414 <NA>  FALSE     FALSE  FALSE FALSE       FALSE   FALSE FALSE
3                   Fencing 1892      1     NA    5.1    81 <NA>  FALSE     FALSE  FALSE FALSE       FALSE   FALSE FALSE
4             Pauvre Pierrot 1892      4     NA    6.7   204 <NA>  FALSE     FALSE  FALSE FALSE       FALSE   FALSE FALSE
5          Blacksmith Scene 1893      1     NA    6.4   679 <NA>  FALSE     FALSE  FALSE FALSE       FALSE   FALSE FALSE
6 Annabelle Butterfly Dance 1894      1     NA    6.1   212 <NA>  FALSE     FALSE  FALSE FALSE       FALSE   FALSE FALSE
```

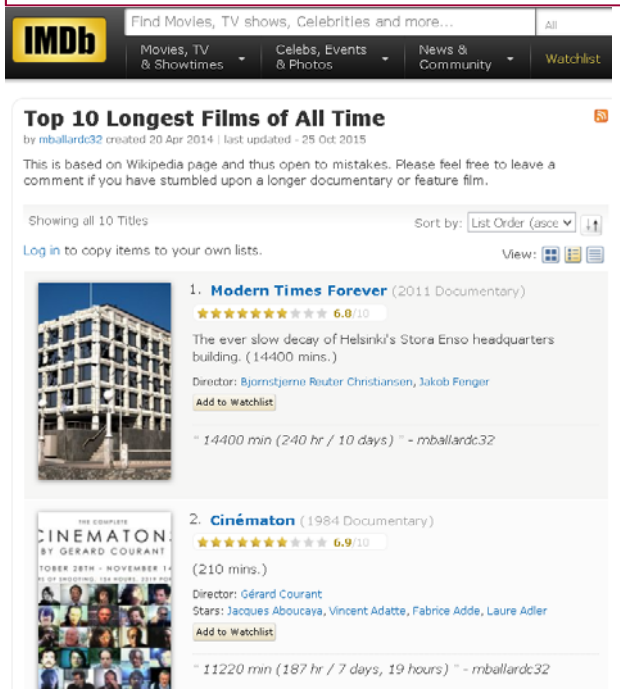
```
> install.packages("gridExtra")
> library(gridExtra)
> g1 <- ggplot(data=movies, aes(length)) + geom_histogram()
> g2 <- ggplot(data=movies, aes(rating)) + geom_histogram() + ylab("votes")
> grid.arrange(g1, g2, nrow=1, ncol=2)
```

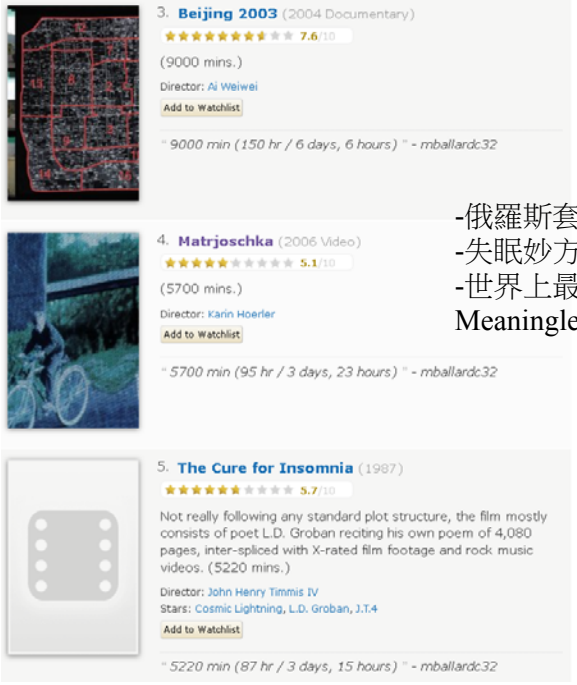


Top 10 longest films

```
> arrange(movies, desc(length))[1:10,] %>% select(title, year, length, rating, votes)
                                                 title year length rating votes
1                                           Matrjoschka 2006   5700    8.5     8
2                                The Cure for Insomnia 1987   5220    5.9   293
3  The Longest Most Meaningless Movie in the World 1970   2880    7.3   143
4                                The Hazards of Helen 1914   1428    6.6    48
5                                                **** 1967   1100    6.9    49
6                                              Resan 1987    873    6.7    40
7                                       Caiyou riji 2008    840    9.2    10
8                          Out 1, noli me tangere 1971    773    7.7   201
9                               Daii jan Napelon 1976    770    7.3   338
10                                 Broken Saints 2003    720    7.5   359
```

See also: 史上超長超無聊電影挑戰人類的耐心極限！
https://read01.com/j6eL0z.html





-俄羅斯套 Matrjoschka 5700分鐘/3天23小時
-失眠妙方 The Cure for Insomnia 5220分鐘/3天15小時
-世界上最長最沒意義的電影The Longest Most Meaningless Movie in the World 2880分鐘/2天

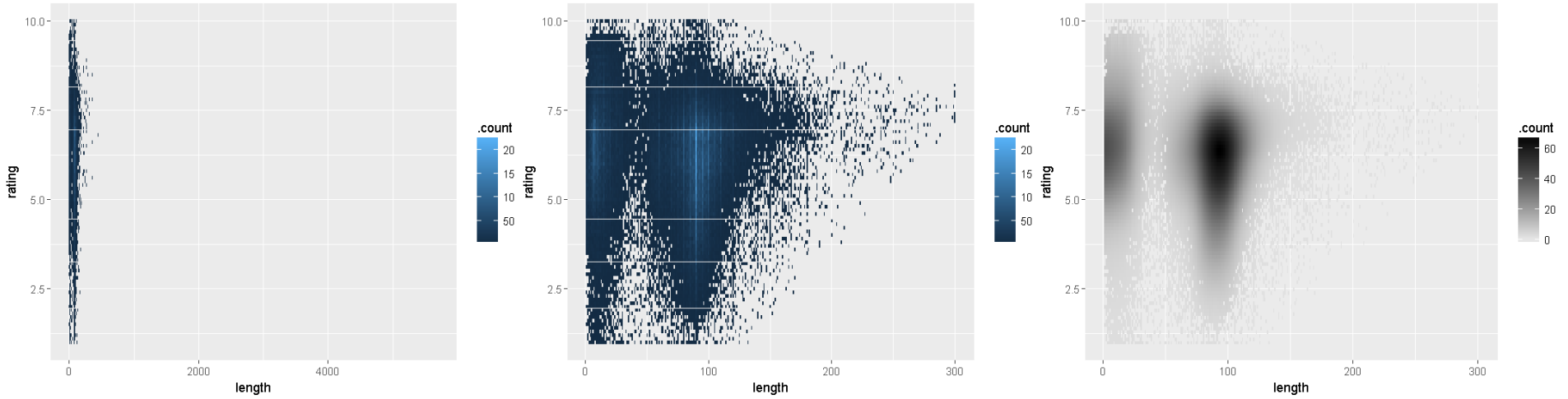
bigvis plot with outliers removed

```
> nobin <- 1e4
> binData <- with(movies, condense(bin(length, find_width(length, nobin)),
+                                  bin(rating, find_width(rating, nobin))))
Summarising with count
> ggplot(data=binData, aes(length, rating, fill = .count)) + geom_tile()
> last_plot() %+% peel(binData)
> smoothBinData <- smooth(peel(binData), h=c(20, 1))
> autoplot(smoothBinData)
```

peel {bigvis}: Peel off low density regions of the data.

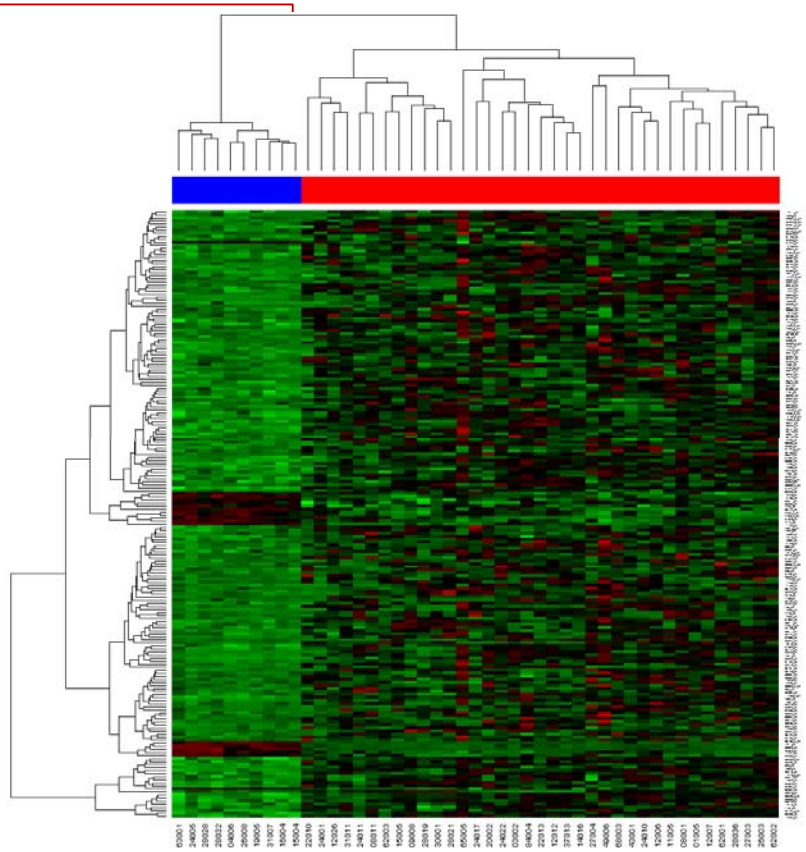
Description: Keeps specified proportion of data by removing the lowest density regions, either anywhere on the plot, or for 2d, just around the edges.

Usage: `peel(x, keep = 0.99, central = NULL)`



heatmap {stats}

```
> source("https://bioconductor.org/biocLite.R")
> biocLite("ALL")
> library(ALL)
> data(ALL)
> ALL
> str(ALL)
> dim(exprs(ALL))
[1] 12625   128
> exprs(ALL)[1:3, 1:5]
             01005    01010    03002    04006    04007
1000_at   7.597323 7.479445 7.567593 7.384684 7.905312
1001_at   5.046194 4.932537 4.799294 4.922627 4.844565
1002_f_at 3.900466 4.208155 3.886169 4.206798 3.416923
> table(ALL$mol.biol)

ALL1/AF4  BCR/ABL E2A/PBX1      NEG   NUP-98  p15/p16
      10       37        5       74        1        1
> eset <- ALL[, ALL$mol.biol %in%
                c("BCR/ABL", "ALL1/AF4")]
> dim(exprs(eset))
[1] 12625    47
> f <- factor(as.character(eset$mol.biol))
> eset.p <- apply(exprs(eset), 1, function(x) t.test(x ~ f)$p.value)
> selected.eset <- eset[eset.p < 0.00001, ]
> dim(selected.eset)
Features  Samples
     200       47
> ma.col <- colorRampPalette(c("green", "black", "red"))(200)
> var.col <- ifelse(f=="ALL1/AF4", "blue", "red")
> heatmap(exprs(selected.eset), col=ma.col, ColSideColors=var.col,
          scale="row")
```

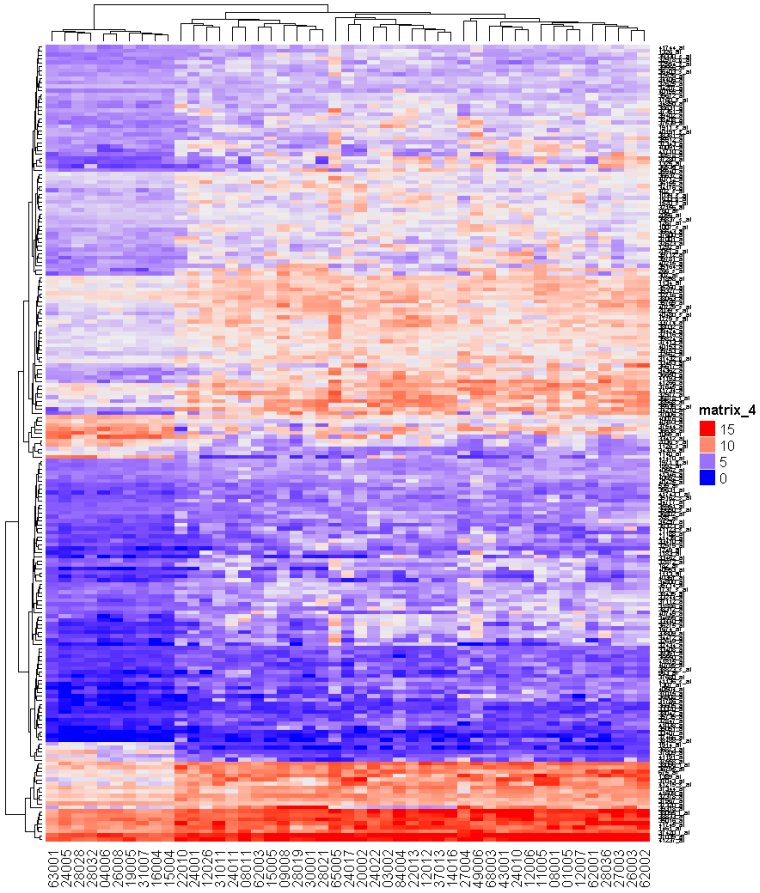


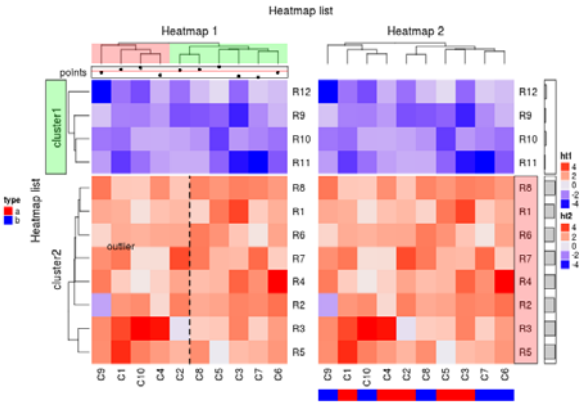
Complex Heatmap

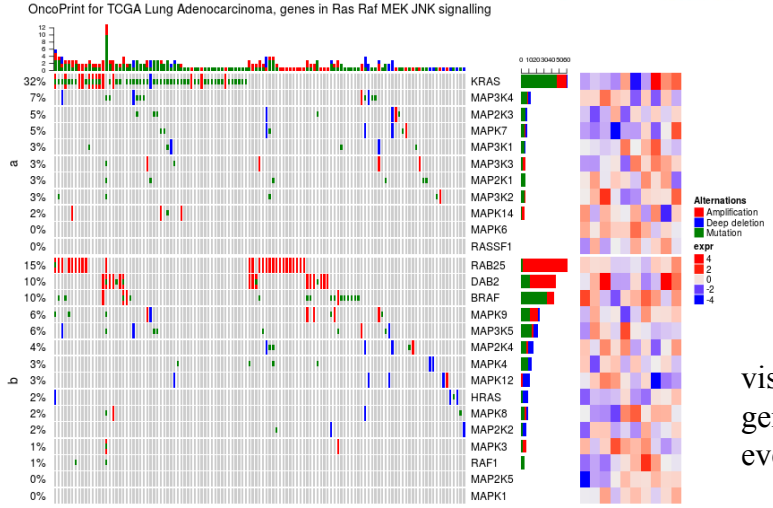
http://www.bioconductor.org/packages/devel/bioc/html/ComplexHeatmap.html

```
> source("https://bioconductor.org/biocLite.R")
> biocLite("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

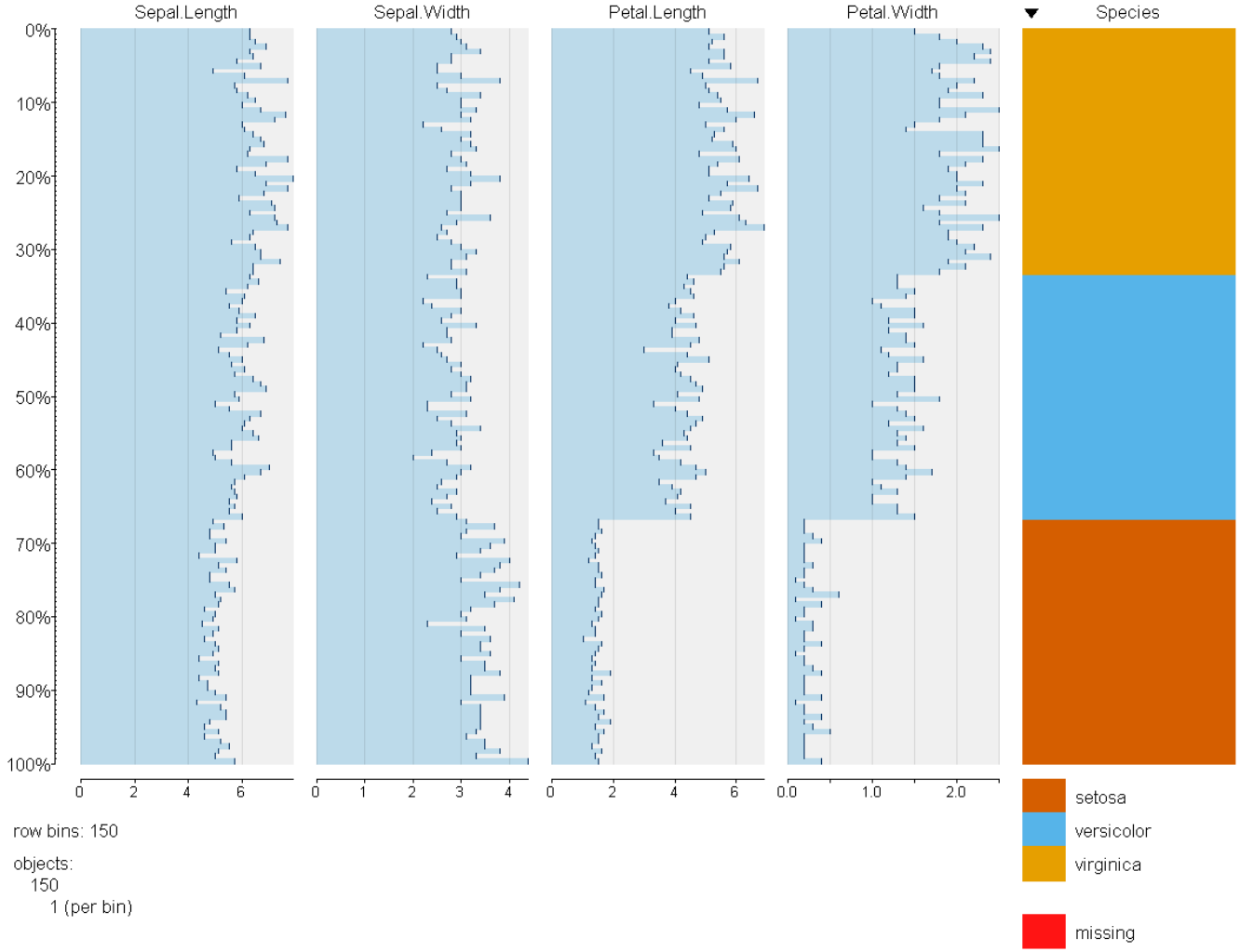
tabplot: Tableplot, a Visualization of Large Datasets

- A tableplot is a visualisation of a (large) dataset with a dozen of variables, both numeric and categorical.
  - Each column represents a variable and each row bin is an aggregate of a certain number of records.
  - Numeric variables are visualized as bar charts, and
  - categorical variables as stacked bar charts. Missing values are taken into account.
  - Also supports large '**ffdf**' datasets from the '**ff**' package.
  - https://github.com/mtennekes/tabplot
  - https://cran.r-project.org/web/packages/tabplot/vignettes/tabplot-vignette.html
- Tennekes, M., Jonge, E. de, Daas, P.J.H. (2013) Visualizing and Inspecting Large Datasets with Tableplots, Journal of Data Science 11 (1), 43-58.

```
tableplot(dat, select, subset = NULL, sortCol = 1, decreasing = TRUE,
  nBins = 100, from = 0, to = 100, nCols = ncol(dat), sample = FALSE,
  sampleBinSize = 1000, scales = "auto", numMode = "mb-sdb-ml",
  max_levels = 50, pals = list("Set1", "Set2", "Set3", "Set4"),
  change_palette_type_at = 20, rev_legend = FALSE, colorNA = "#FF1414",
  colorNA_num = "gray75", numPals = "OrBu", limitsX = NULL,
  bias_brokenX = 0.8, IQR_bias = 5, select_string = NULL,
  subset_string = NULL, colNames = NULL, filter = NULL, plot = TRUE,
  ...)
```

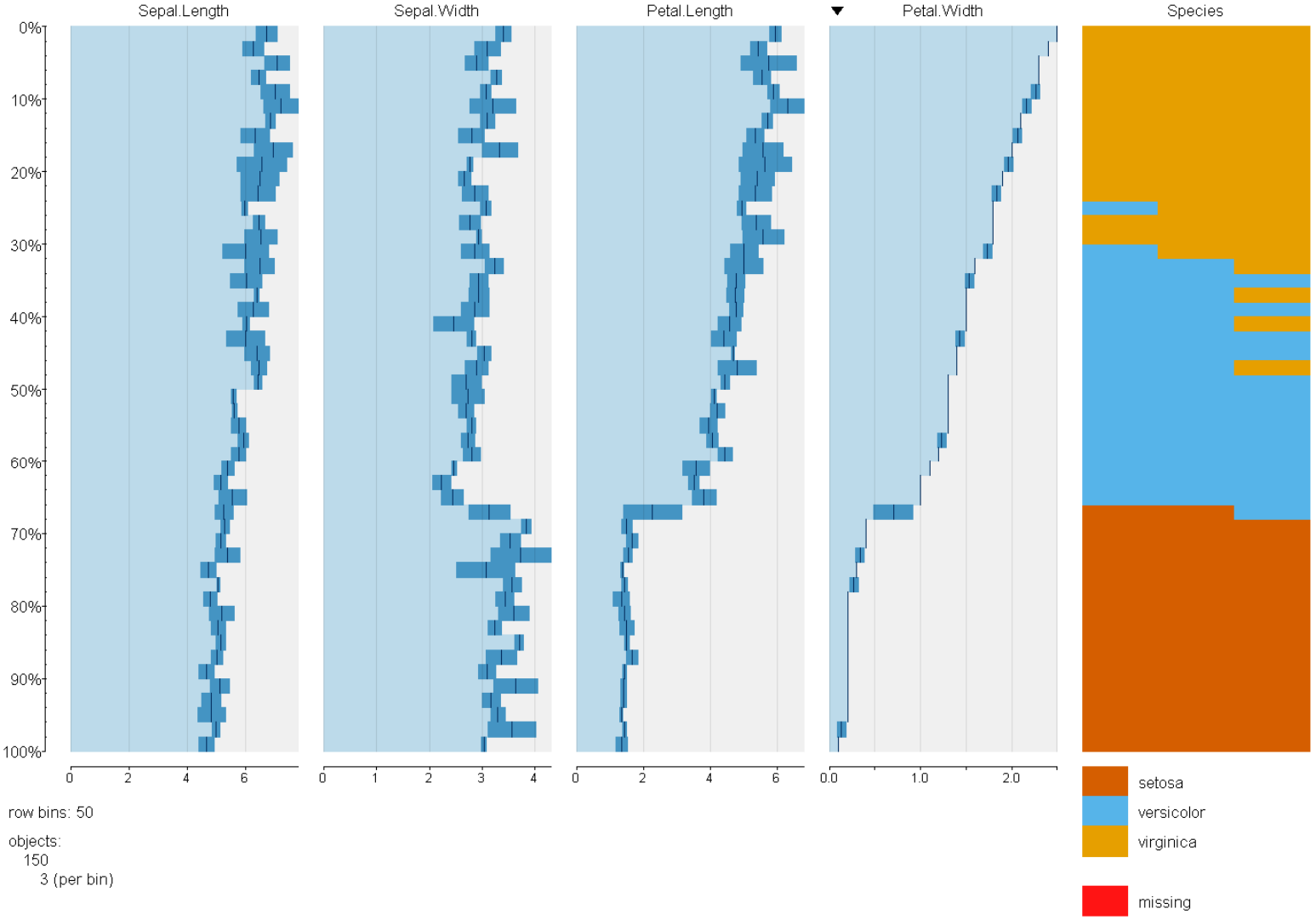
tableplot(iris, nBins=150, sortCol=5)

```
> install.packages("tabplot")
> library(tabplot)
> tableplot(iris, nBins=150, sortCol=5)
```



tableplot(iris, nBins=50, sortCol=4)

```
> tableplot(iris, nBins=50, sortCol=4)
```

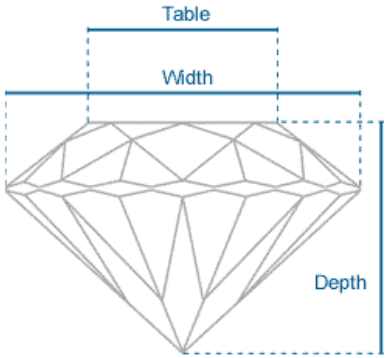


tableplot(diamonds)

```
> require(ggplot2)
> data(diamonds)
> dim(diamonds)
[1] 53940    10
> head(diamonds)
# A tibble: 6 × 10
  carat       cut color clarity depth table price     x     y     z
  <dbl>     <ord> <ord>   <ord> <dbl> <dbl> <int> <dbl> <dbl> <dbl>
1  0.23     Ideal     E     SI2  61.5    55   326  3.95  3.98  2.43
2  0.21   Premium     E     SI1  59.8    61   326  3.89  3.84  2.31
3  0.23      Good     E     VS1  56.9    65   327  4.05  4.07  2.31
4  0.29   Premium     I     VS2  62.4    58   334  4.20  4.23  2.63
5  0.31      Good     J     SI2  63.3    58   335  4.34  4.35  2.75
6  0.24 Very Good     J    VVS2  62.8    57   336  3.94  3.96  2.48
> tableplot(diamonds)
```

Details

- price. price in US dollars (\$326--\$18,823)
- carat. weight of the diamond (0.2--5.01)
- cut. quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- colour. diamond colour, from J (worst) to D (best)
- clarity. a measurement of how clear the diamond is (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best))
- x. length in mm (0--10.74)
- y. width in mm (0--58.9)
- z. depth in mm (0--31.8)
- depth. total depth percentage = z / mean(x, y) = 2 * z / (x + y) (43--79)
- table. width of top of diamond relative to widest point (43--95)

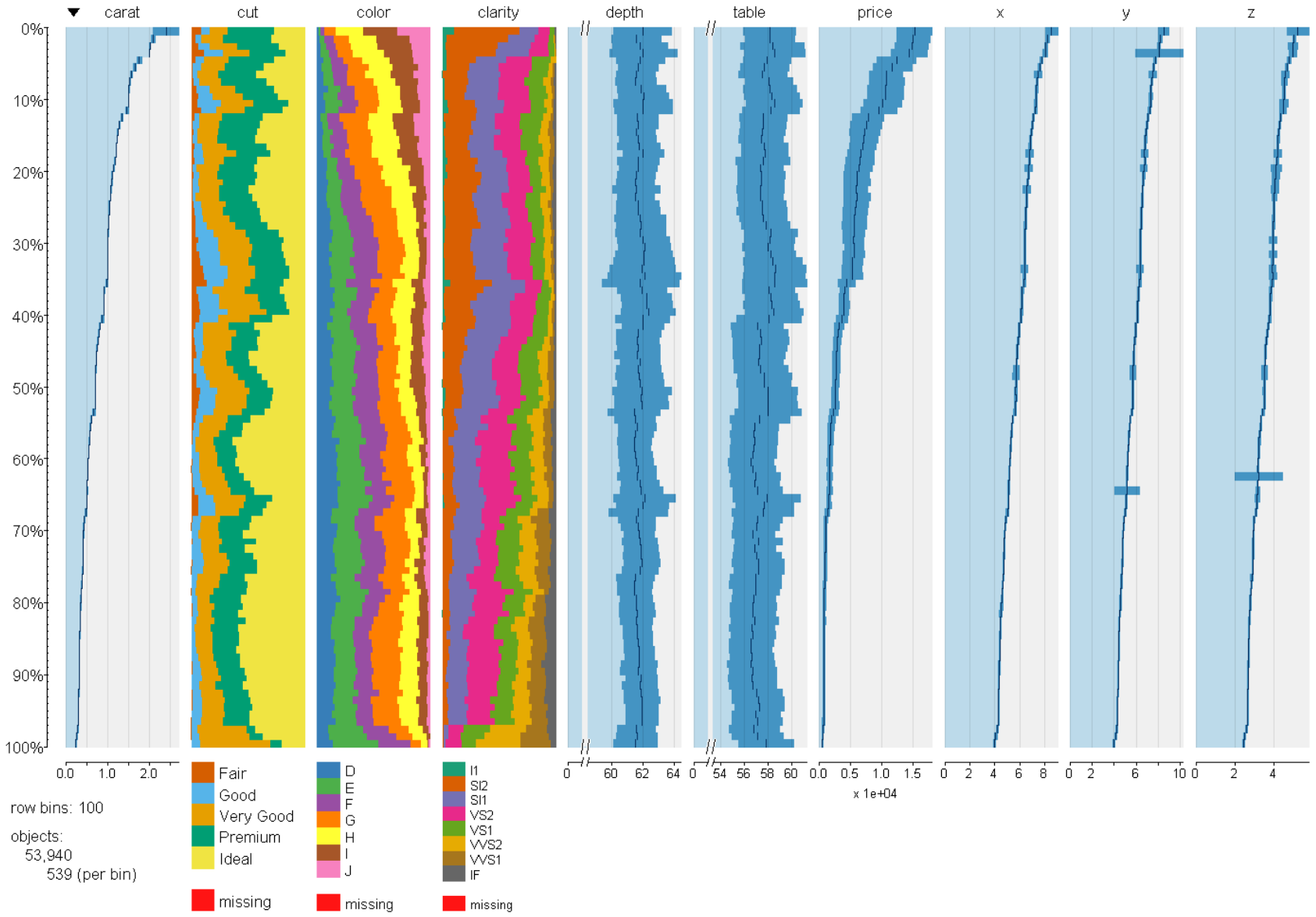


http://www.lumeradiamonds.com/diamond-education/diamond-cut

http://docs.ggplot2.org/0.9.3.1/diamonds.html

http://yourdiamondteacher.com/diamond-4cs/cut/

tableplot(diamonds)



ggplot2.SparkR: Rebooting ggplot2 for Scalable Big Data Visualization

ggplot2.SparkR

Welcome to ggplot2.SparkR

View project on GitHub

Overview

ggplot2.SparkR is an R package for scalable visualization of big data represented in Spark DataFrame.

It is an extension to the original ggplot2 package and can seamlessly handle both R data.frame and Spark DataFrame with no modifications to the original API.

Installation

SparkR Installation

Build Spark

Build Spark with Maven and include the `-PsparkR` profile to build the R package. For example to use the default Hadoop versions you can run

```
build/mvn -DskipTests -Psparkr package
```

Using SparkR from RStudio

If you wish to use SparkR from RStudio or other R frontends you will

Download .zip file

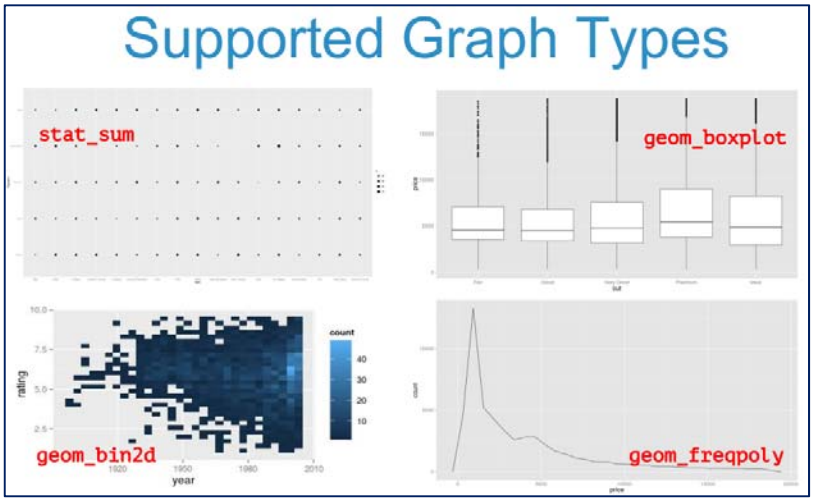
Download .tar.gz file

is maintained by SKKU-SKT.

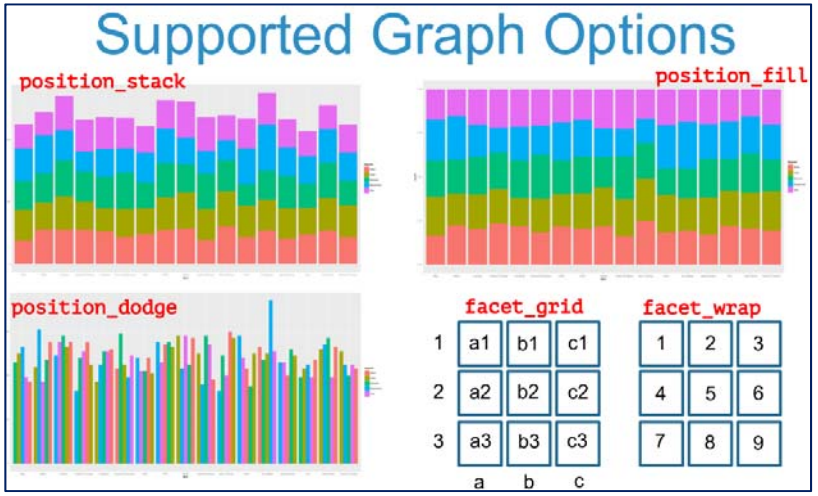
This page was generated by GitHub Pages using the Architect theme by Jason Long.

- Home
- Hello ggplot2.SparkR
- Supported Plot Types

Supported Graph Types

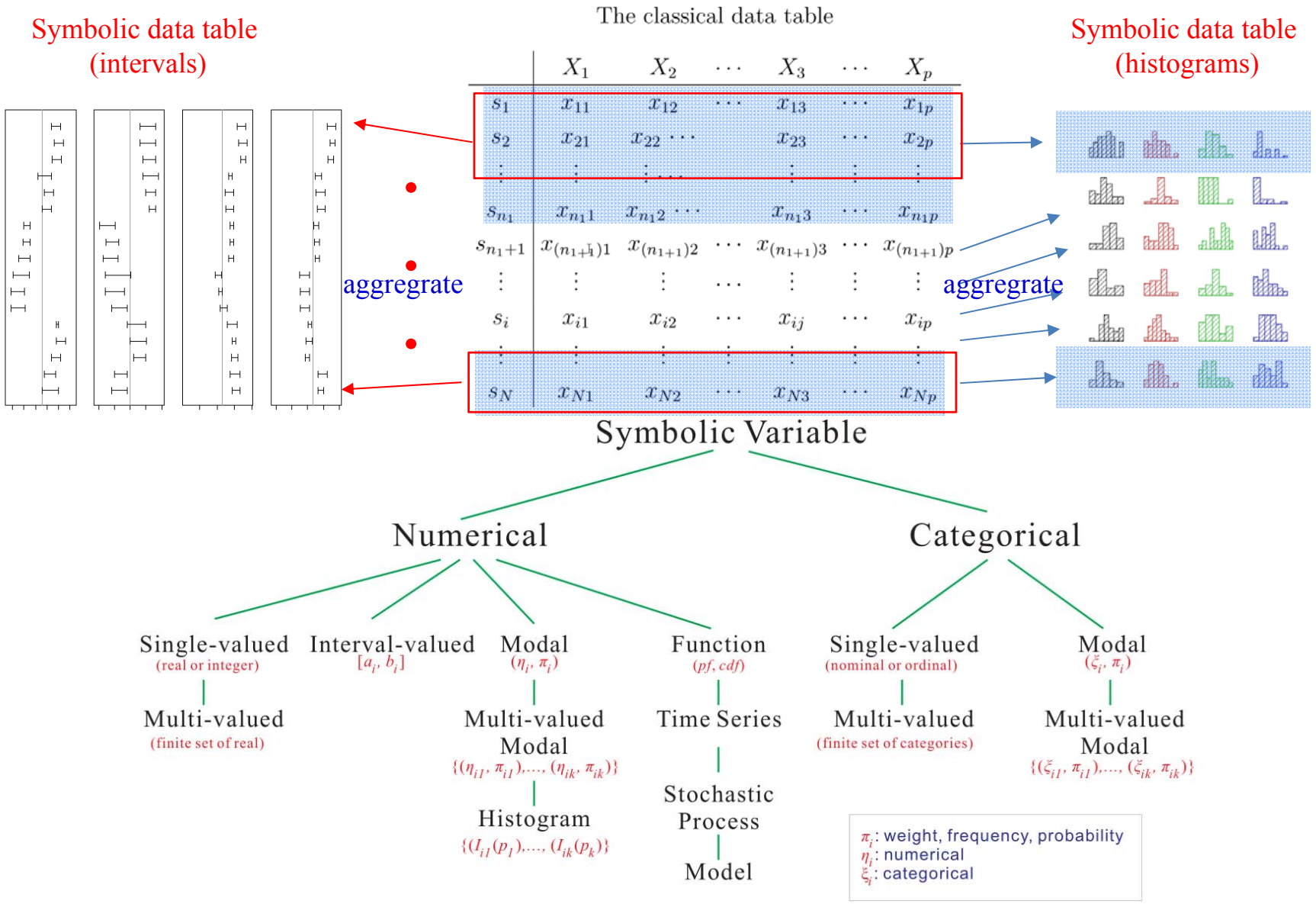


Supported Graph Options



http://skku-skt.github.io/ggplot2.SparkR/

Symbolic Data Analysis (Billard and Diday, JASA 2003)



graphics.SDA

the R base graphics package companion for symbolic data analysis

The R Graphics Package Companion for Symbolic Data Analysis

Documentation for package 'graphics.SDA' version 0.0.0.9

- DESCRIPTION file.
- Code demos. Use demo() to run them.

Help Pages

boxplot.i	Box Plots For Interval Data
boxplot.sbs.i	The Side-by-Side Box Plots For Interval Data
cbind.h	Combine HistData Objects by Rows or Columns
cbind.i	Combine IntervalData Objects by Rows or Columns
get.subset	The Subset of the Histogram Data
hM.size	The Dimensions of the Histogram Data
image.i	The Image Plot For Interval Data
plot.index.i	The Index Plot For Interval Data

Not Yet Released!

plot.index.i {graphics.SDA}

The Index Plot For Interval Data

Description

The index plot for one sample of the interval data

Usage

```
## S3 method for class 'index.i'
plot(idata, vertical = FALSE, type = "seg", align = "d",
  fill.col = "lightcyan", col = "black", cex = 0.85,
  show.mean.value = F, ...)
```

Arguments

idata	one sample of an IntervalData object, or the data matirx with the (min, max) format
vertical	logical.
type	"seg": segments, "rect": rectangular
align	the direction (x-axis or y-axis) of the indics default(d), initial(i), left(l), right(r).
fill.col	the filled color of the interval bars
col	the color of the interval segments
...	additional plotting parameters

Details

...

Author(s)

Han-Ming Wu

See Also

plot.2d.i

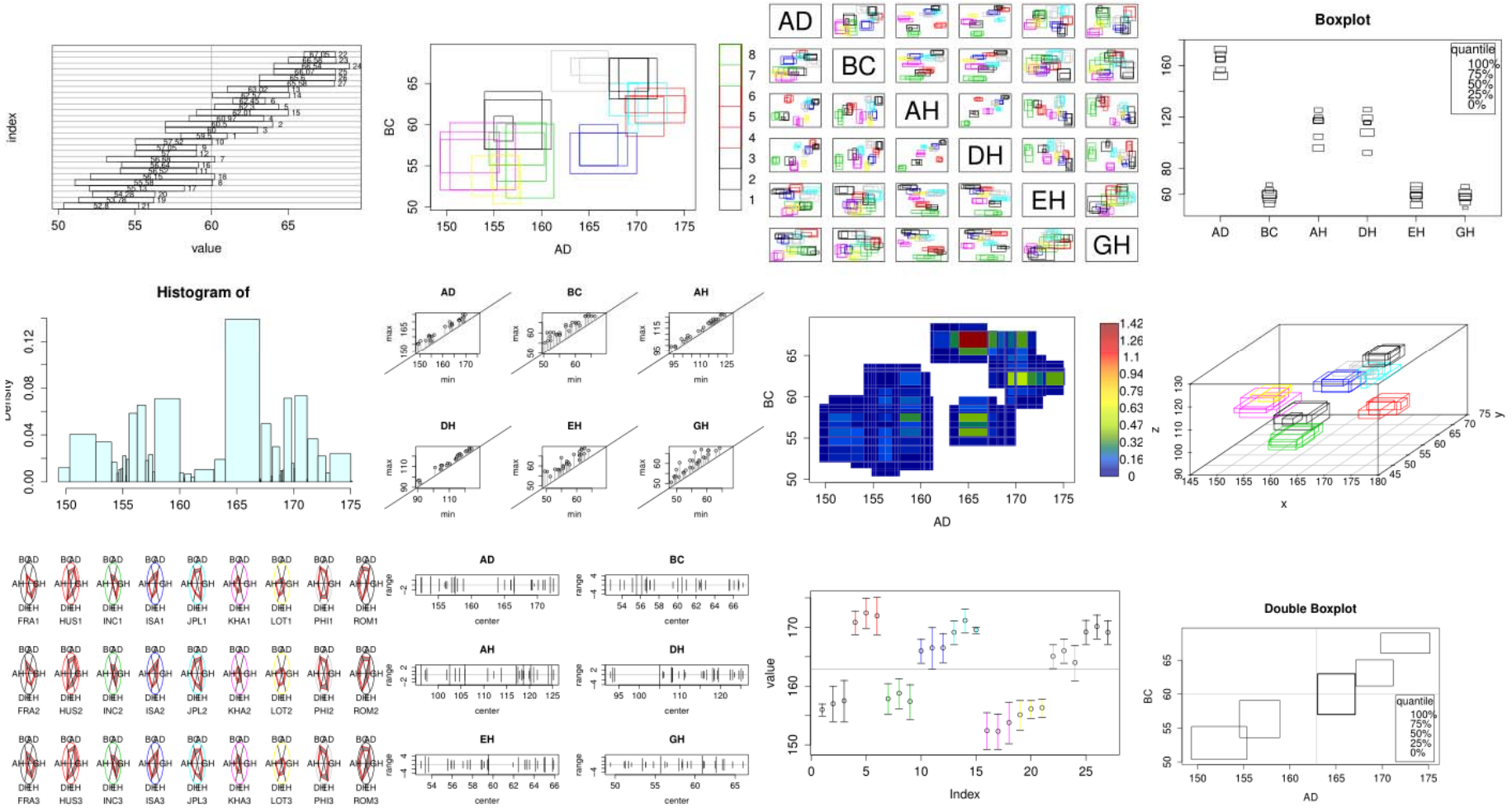
Examples

```
data(face)
idata.x <- face$x
y.C <- face$y
title <- "face data"
plot.index.i(idata.x)
plot.index.i(idata.x, vertical=F)
plot.index.i(idata.x, vertical=F, col=y.C)
plot.index.i(idata.x)
```

graphics.SDA

the R base graphics package companion for symbolic data analysis

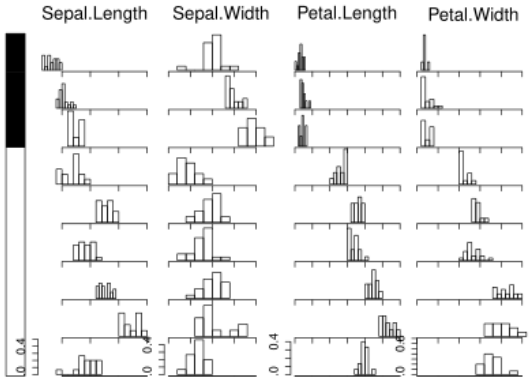
```
plot(interval_object); hist(interval_object); ...
```

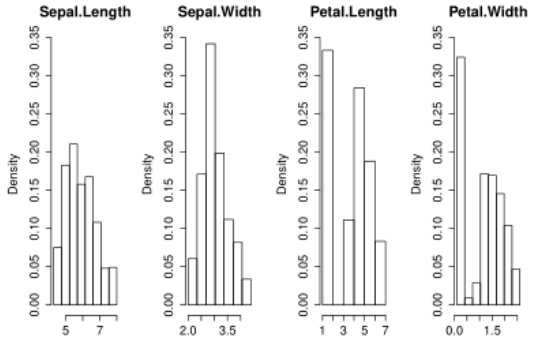


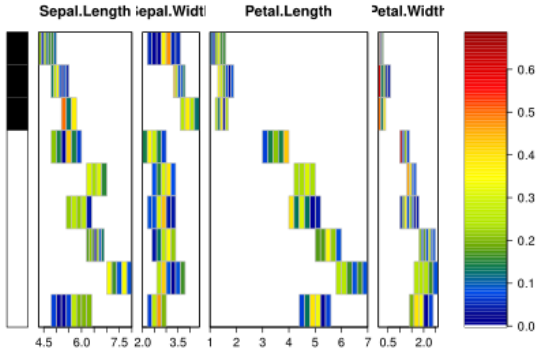
graphics.SDA

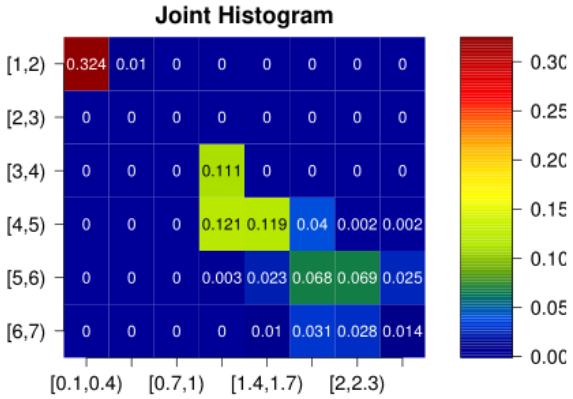
the R base graphics package companion for symbolic data analysis

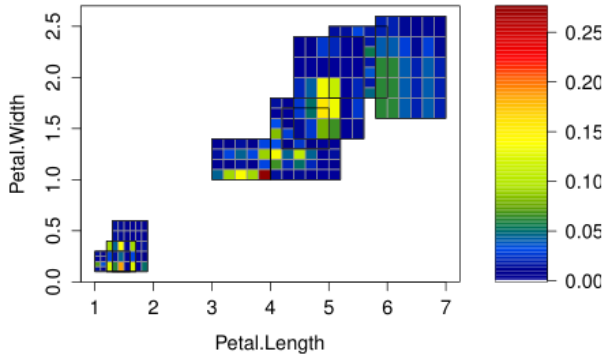
```
plot(histogram_object); hist(histogram_object); ...
```

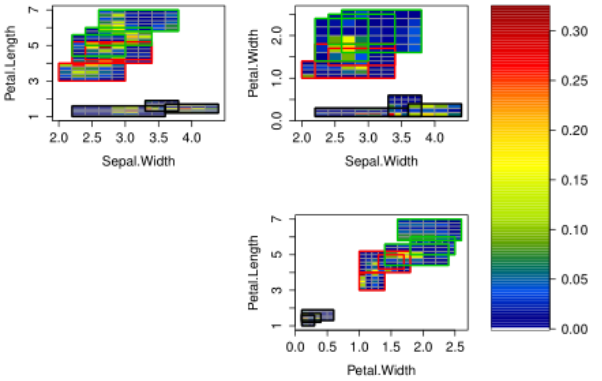






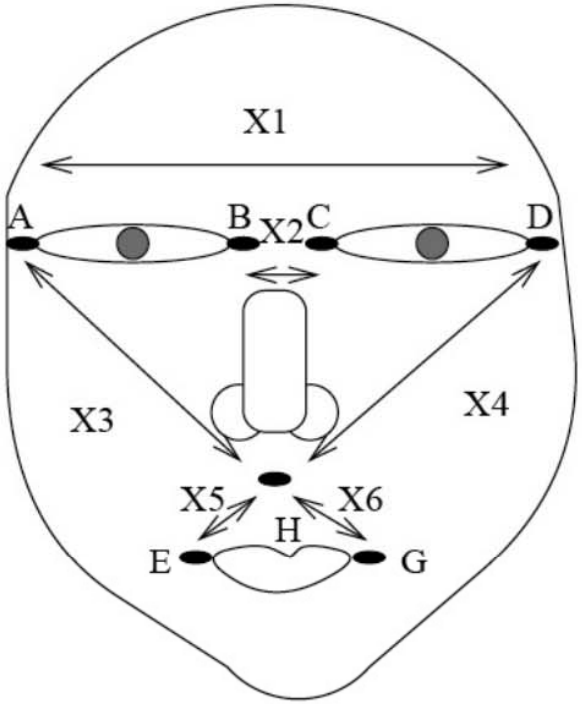






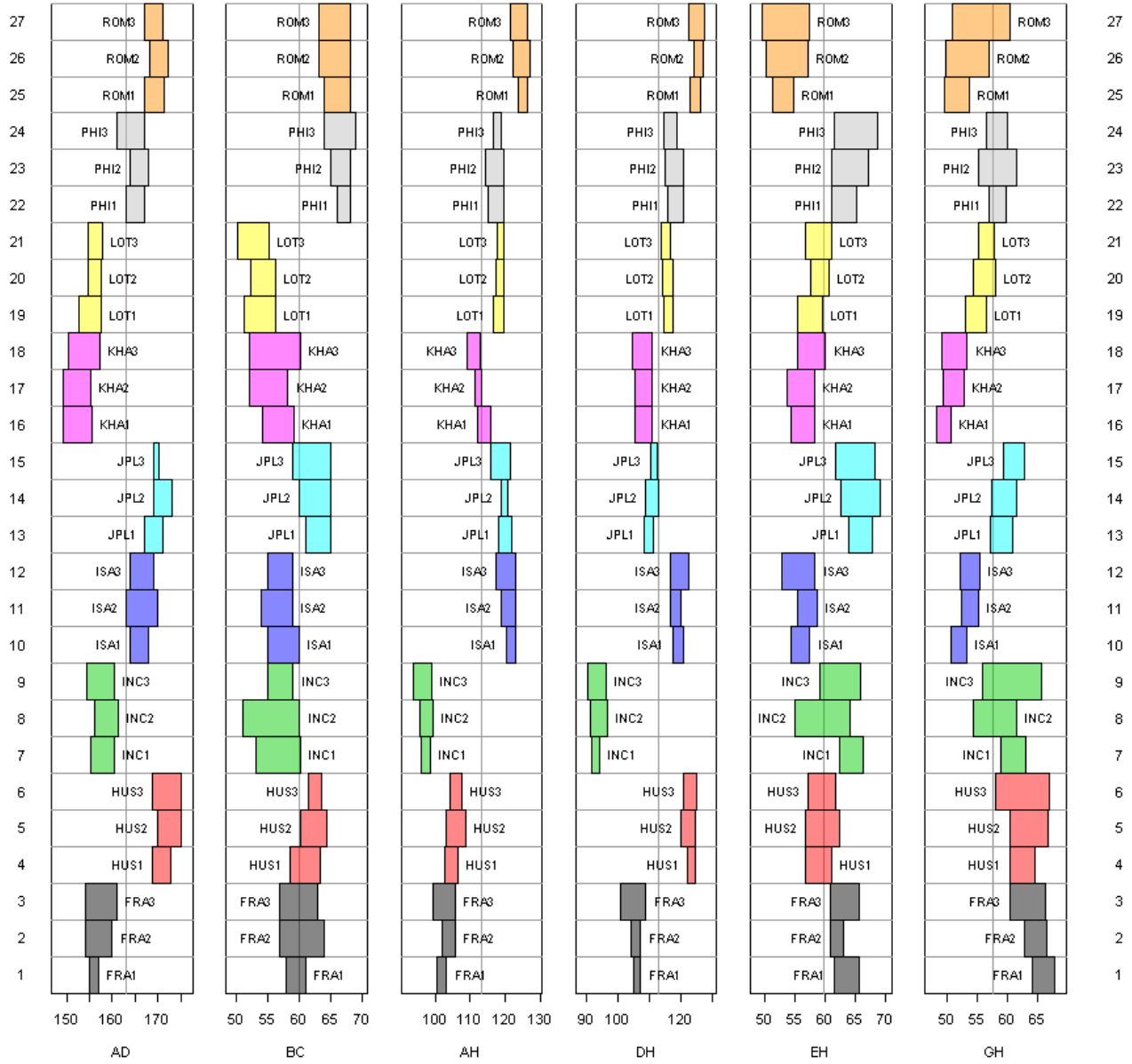
Face Recognition Data

1. Details: Leroy et al. (1996), Douzal-Chouakria, Billard and Diday (2011), Le-Rademacher and Billard (2012)
2. The dataset gives six face measurements of nine men, each with three observations, resulting in a total 27 observations. The measurements for each observation came from a sequence of images.



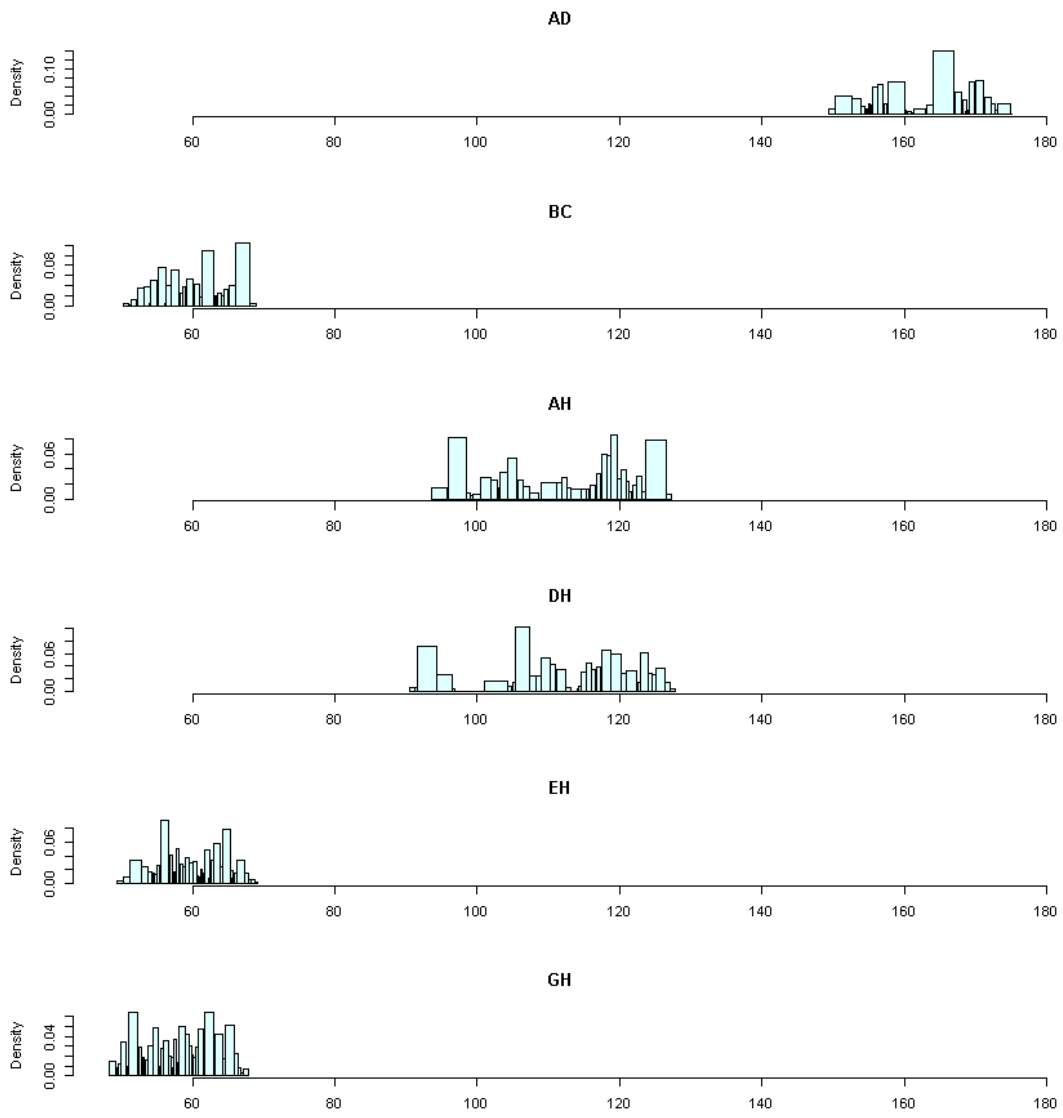
name	AD	BC	AH	DH	EH	GH
FRA1	(155, 157)	(58, 61.01)	(100.45, 103.28)	(105, 107.3)	(61.4, 65.73)	(64.2, 67.8)
FRA2	(154, 160.01)	(57, 64)	(101.98, 105.55)	(104.35, 107.3)	(60.88, 63.03)	(62.94, 66.47)
FRA3	(154.01, 161)	(57, 63)	(99.36, 105.65)	(101.04, 109.04)	(60.95, 65.6)	(60.42, 66.4)
HUS1	(168.86, 172.84)	(58.55, 63.39)	(102.83, 106.53)	(122.38, 124.52)	(56.73, 61.07)	(60.44, 64.54)
⋮			⋮			
ROM3	(167.11, 171.19)	(63.13, 68.03)	(121.62, 126.57)	(122.58, 127.78)	(49.41, 57.28)	(50.99, 60.46)

plot.index.i(idata.x, type="rect", vertical = T, fill.col=y.C)

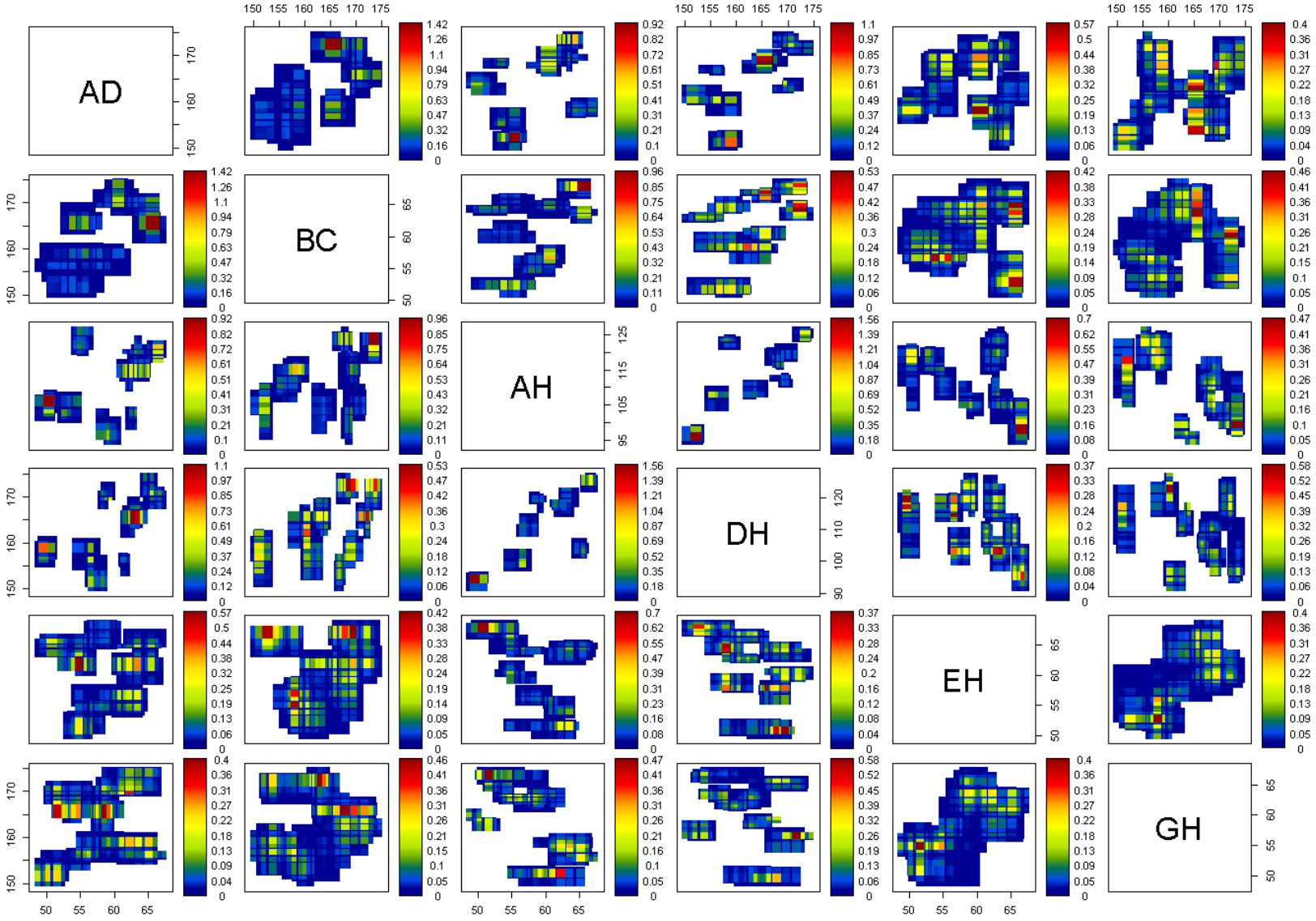


pairs.i(idata.x, col=NA, border.col=y.C)

hist.overlap.i(idata.x)



hist.overlap.2d.i(idata.x, col="rb130")



scatterplot3d.i(idata.x[,1:3,], col=y.C, rgl=T)

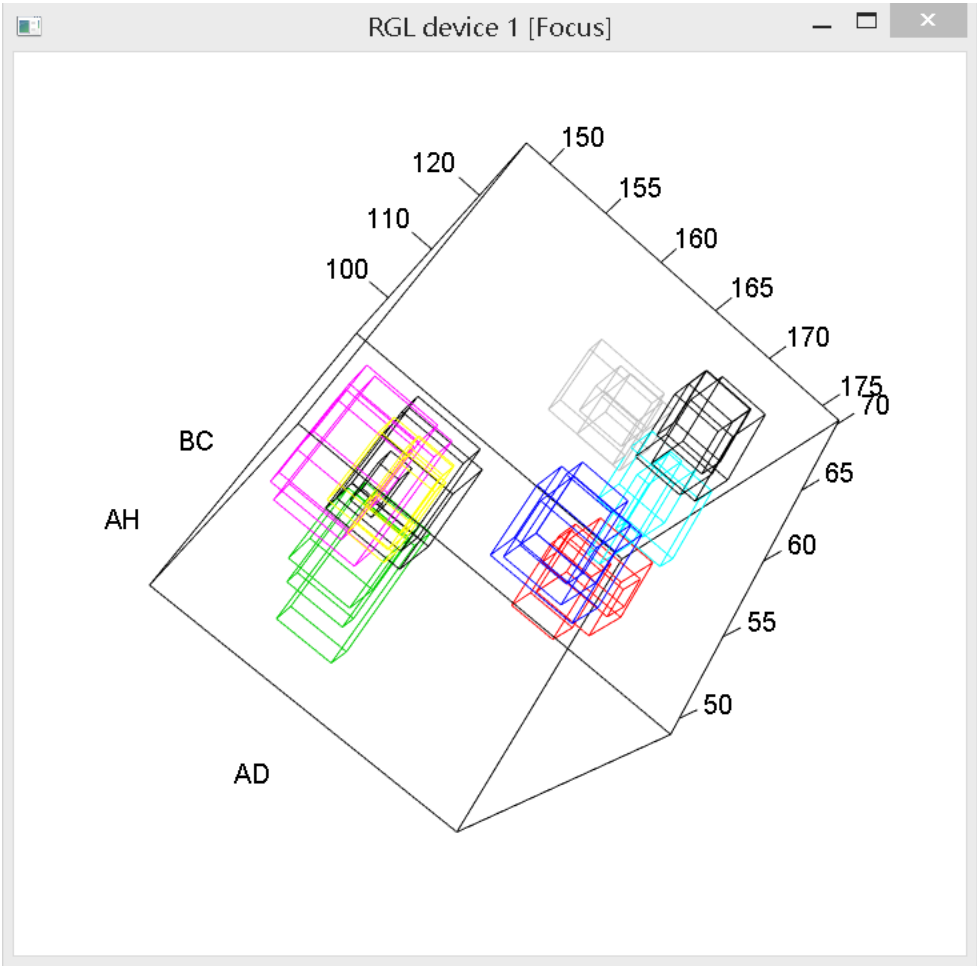
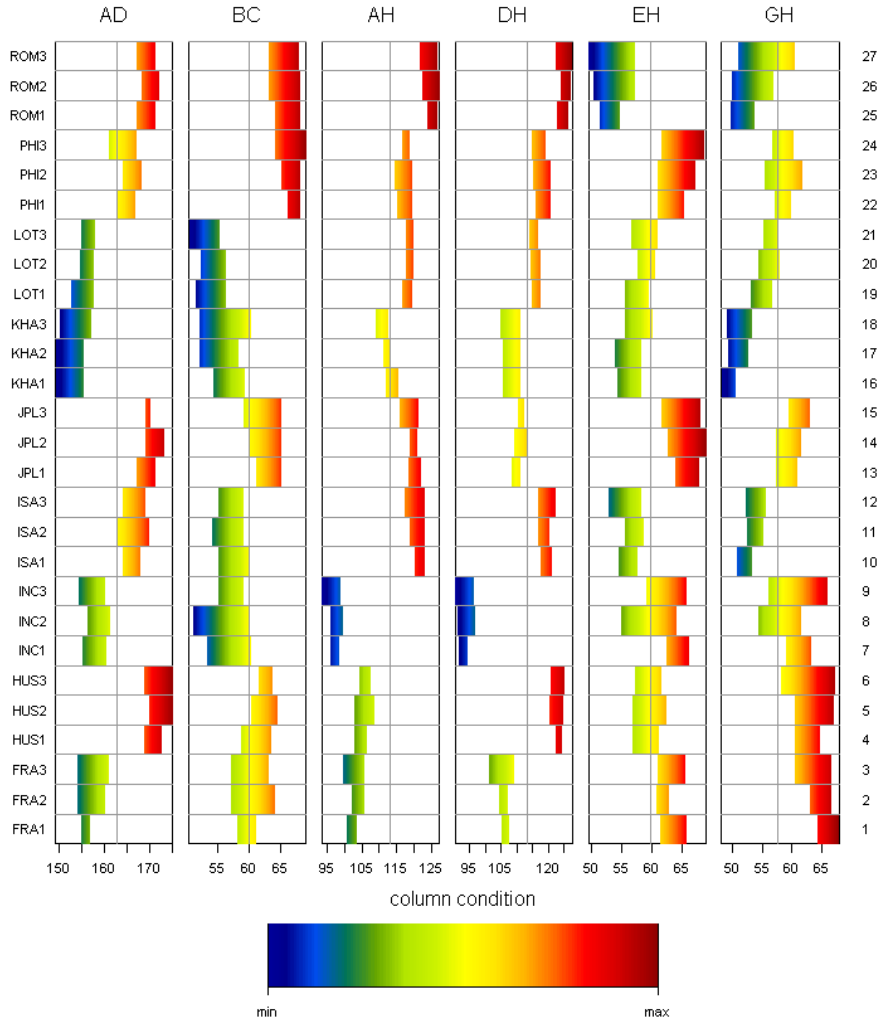
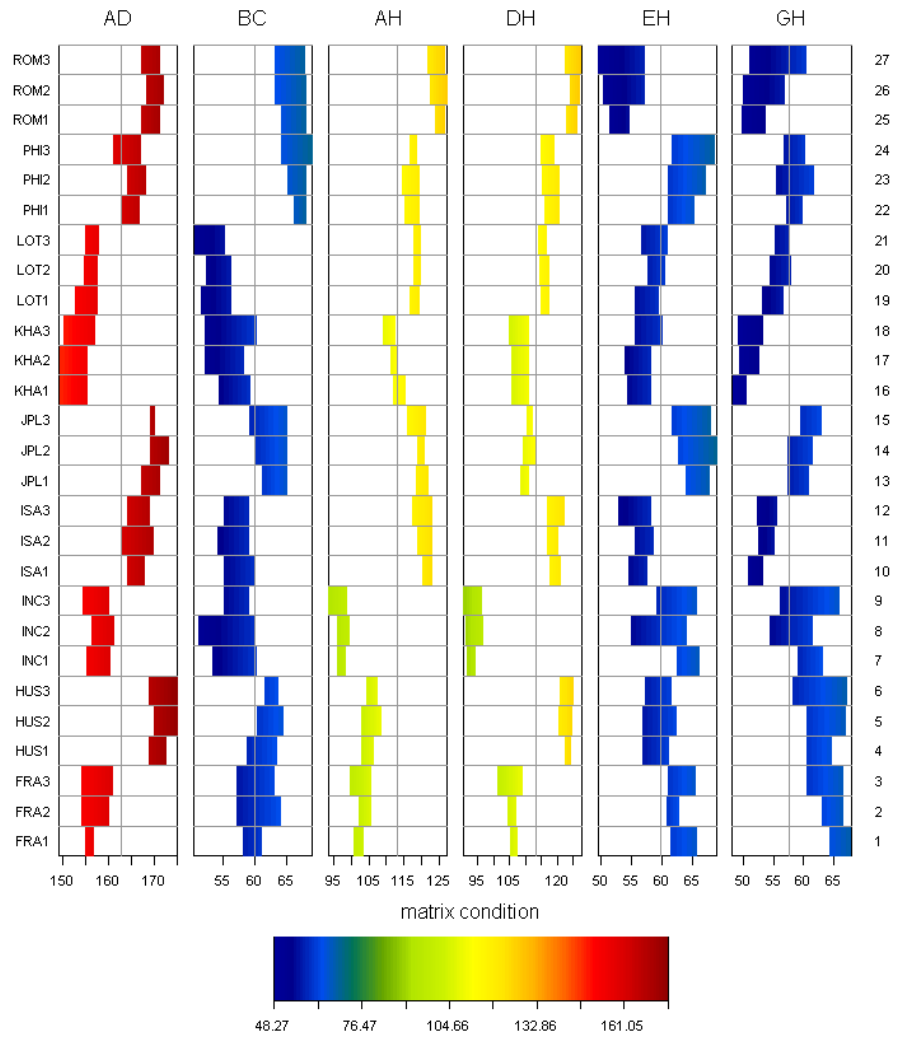
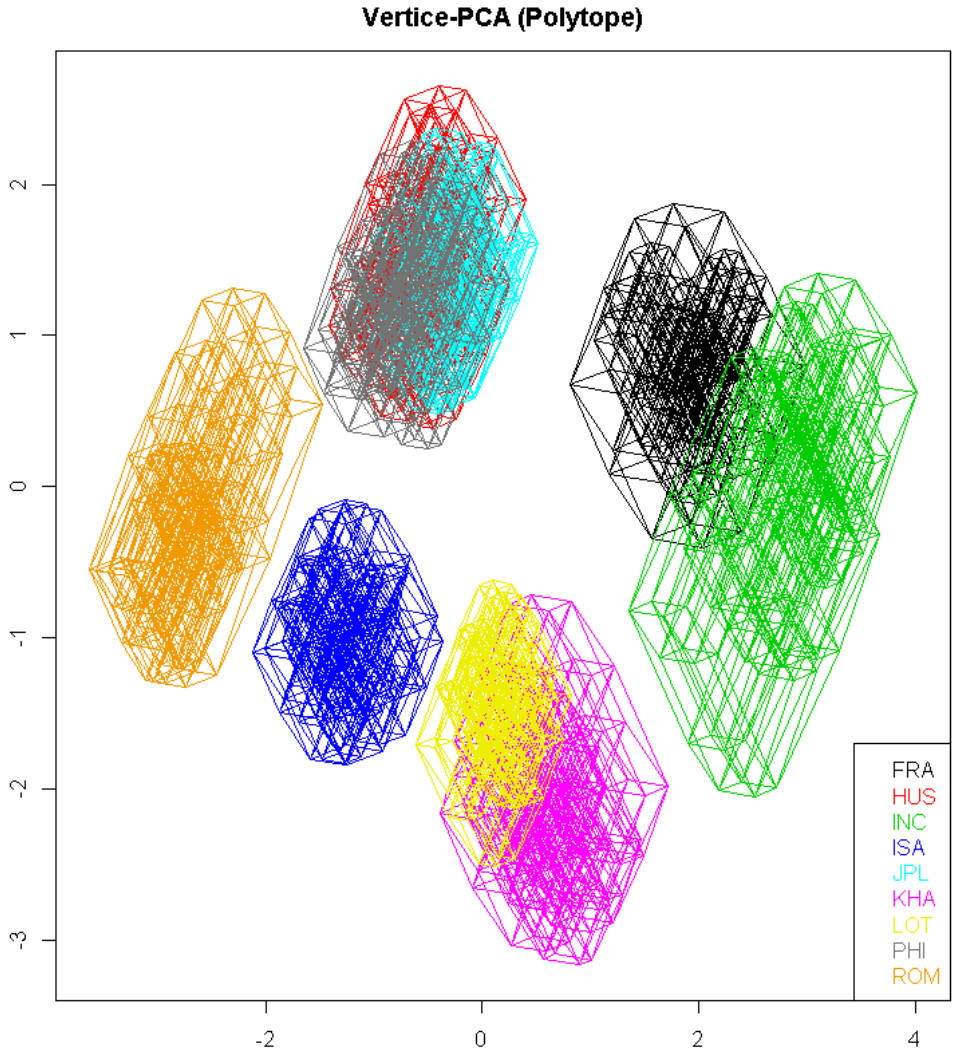


image.i(idata.x)





PCA for interval-valued Data

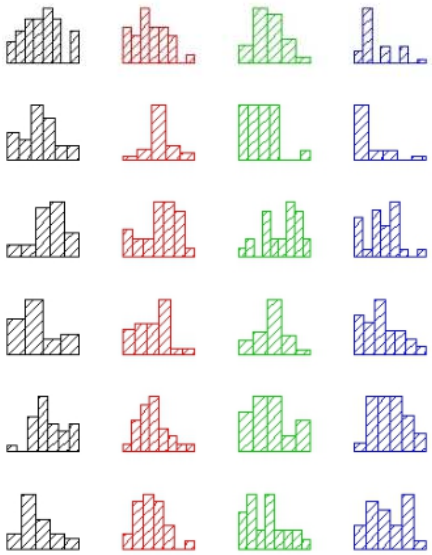


The Histogram-valued data

Let $O = \{o_1, o_2, \cdots, o_n\}$ be given objects.

1. The histogram-valued data: $H = (h_{ij})_{n \times p}$, where

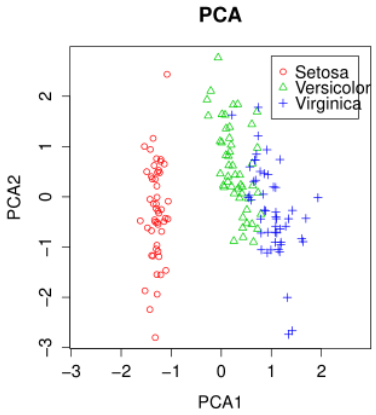
$$h_{ij} = \{I_{ij}, p_{ij}\} = \{x \in I_{ij}^{(b)} = (x_{ij}^{(b)}, x_{ij}^{(b+1)}], p_{ij}^{(b)}, b = 1, \cdots, B_{ij}\}.$$

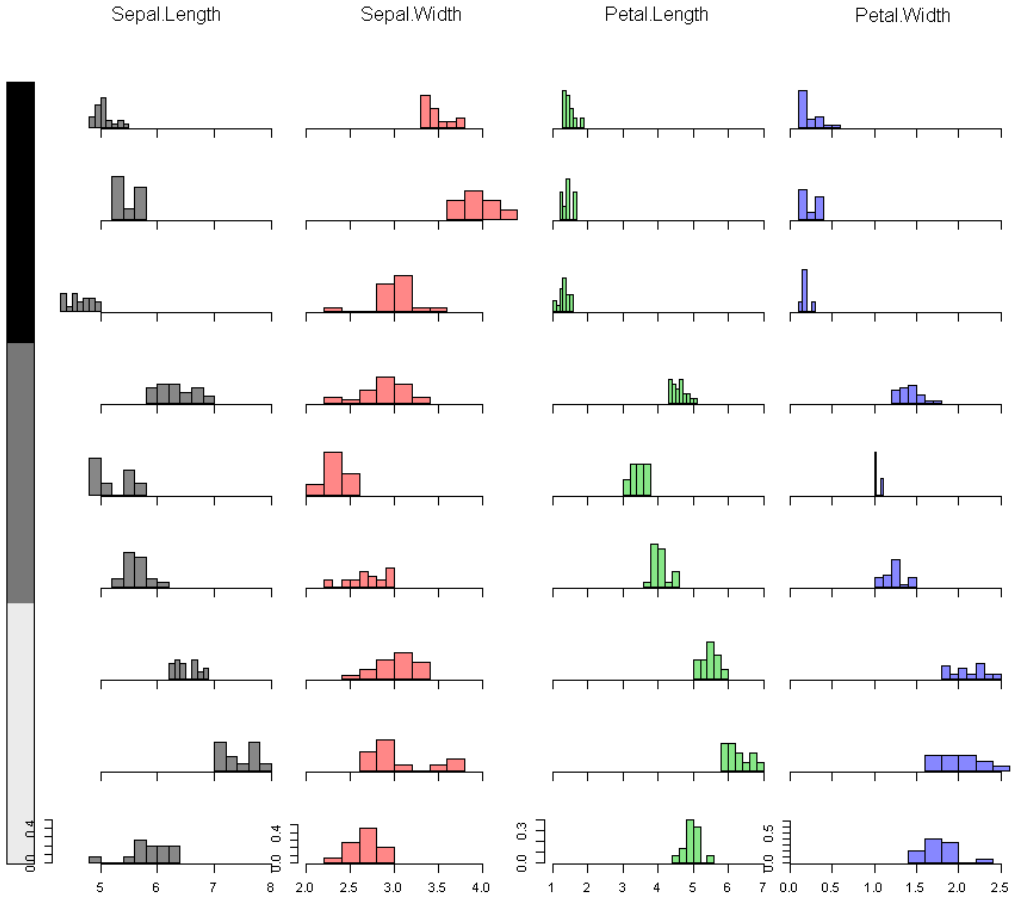


Possible source: the result of an aggregation, the description of a population, or any other grouped collective.

Iris data: generate histograms

The iris data (Fisher, 1936) consists of 50 samples from each of three species of Iris (Setosa, Virginica and Versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres.





hist.h(hdata, scale.x = TRUE)

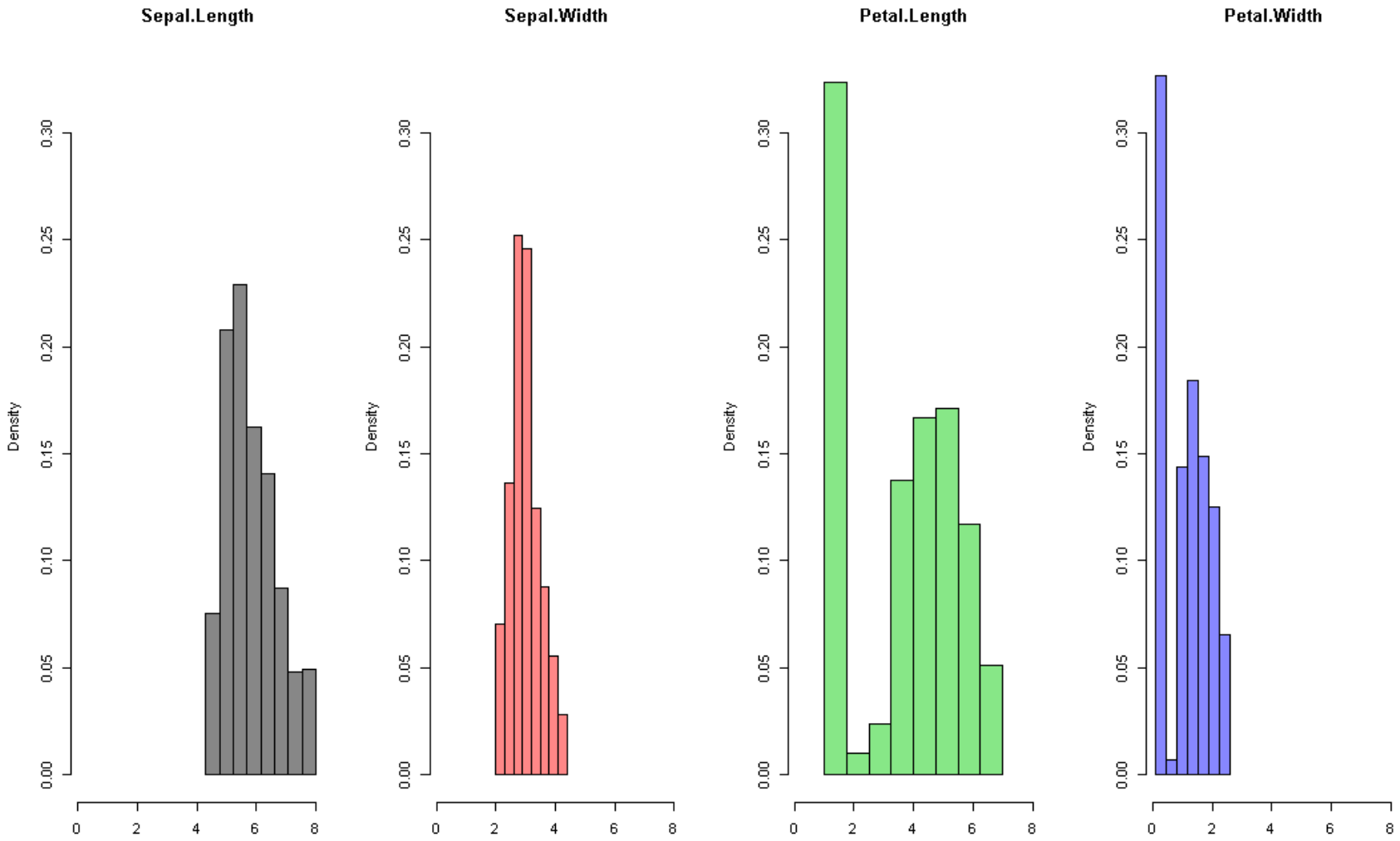
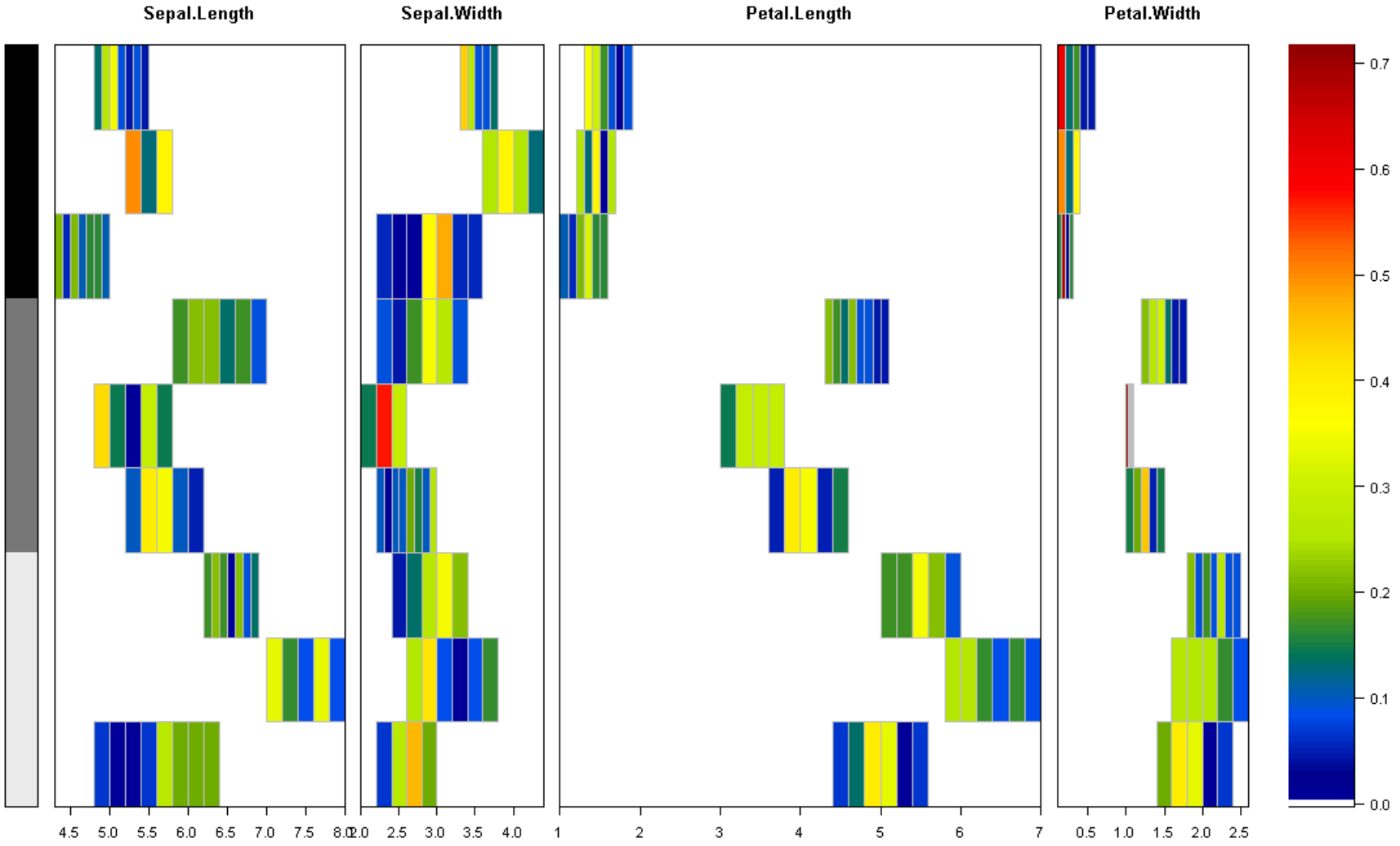
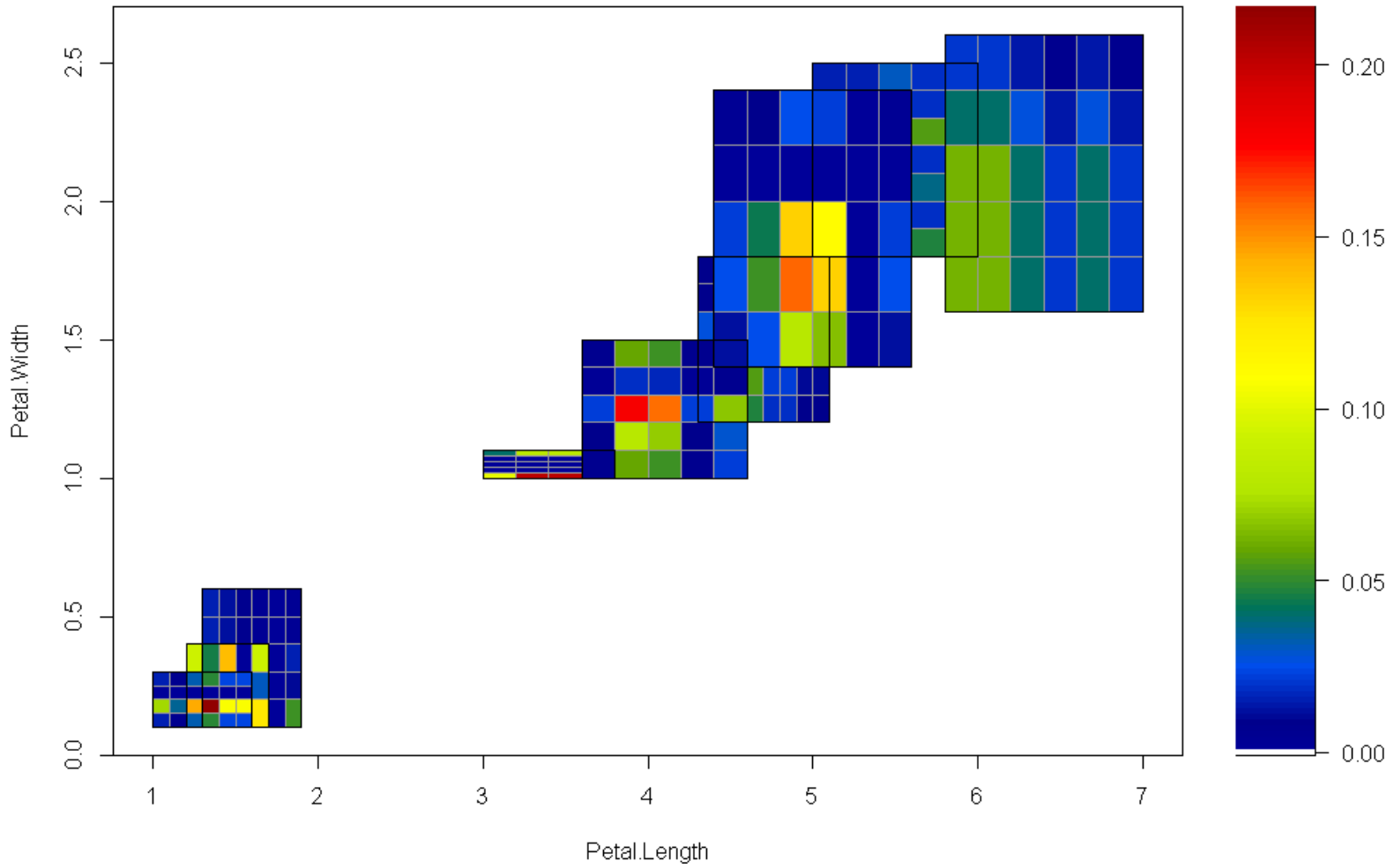


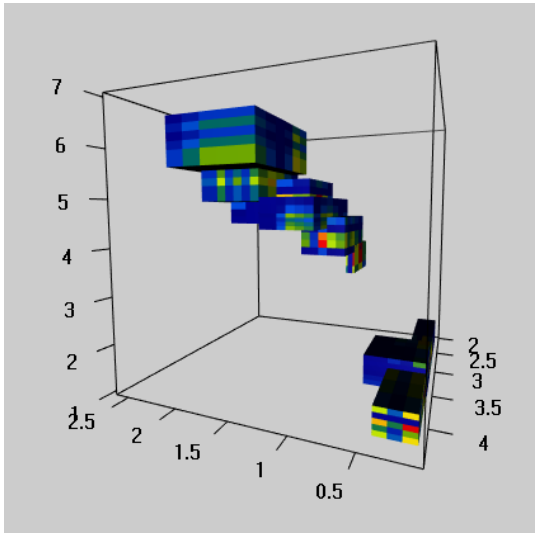
image.h(hdata)

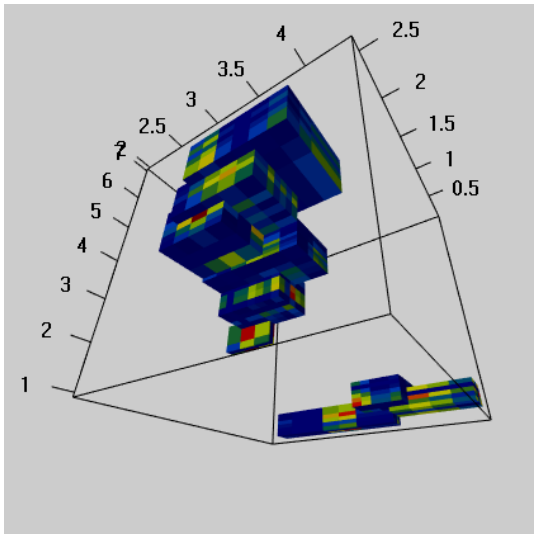


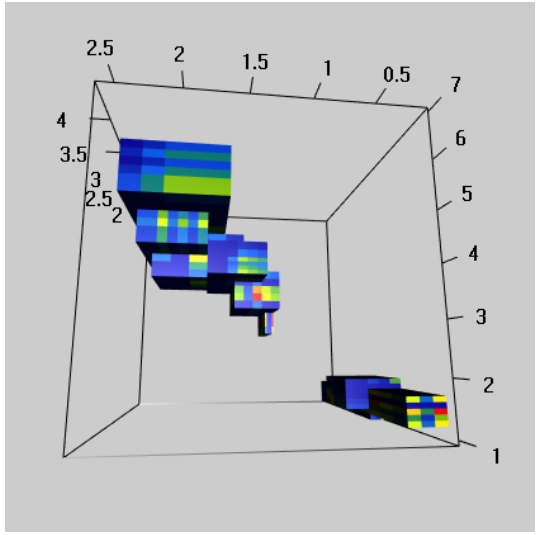
scatterplot.h(hdata.s1)



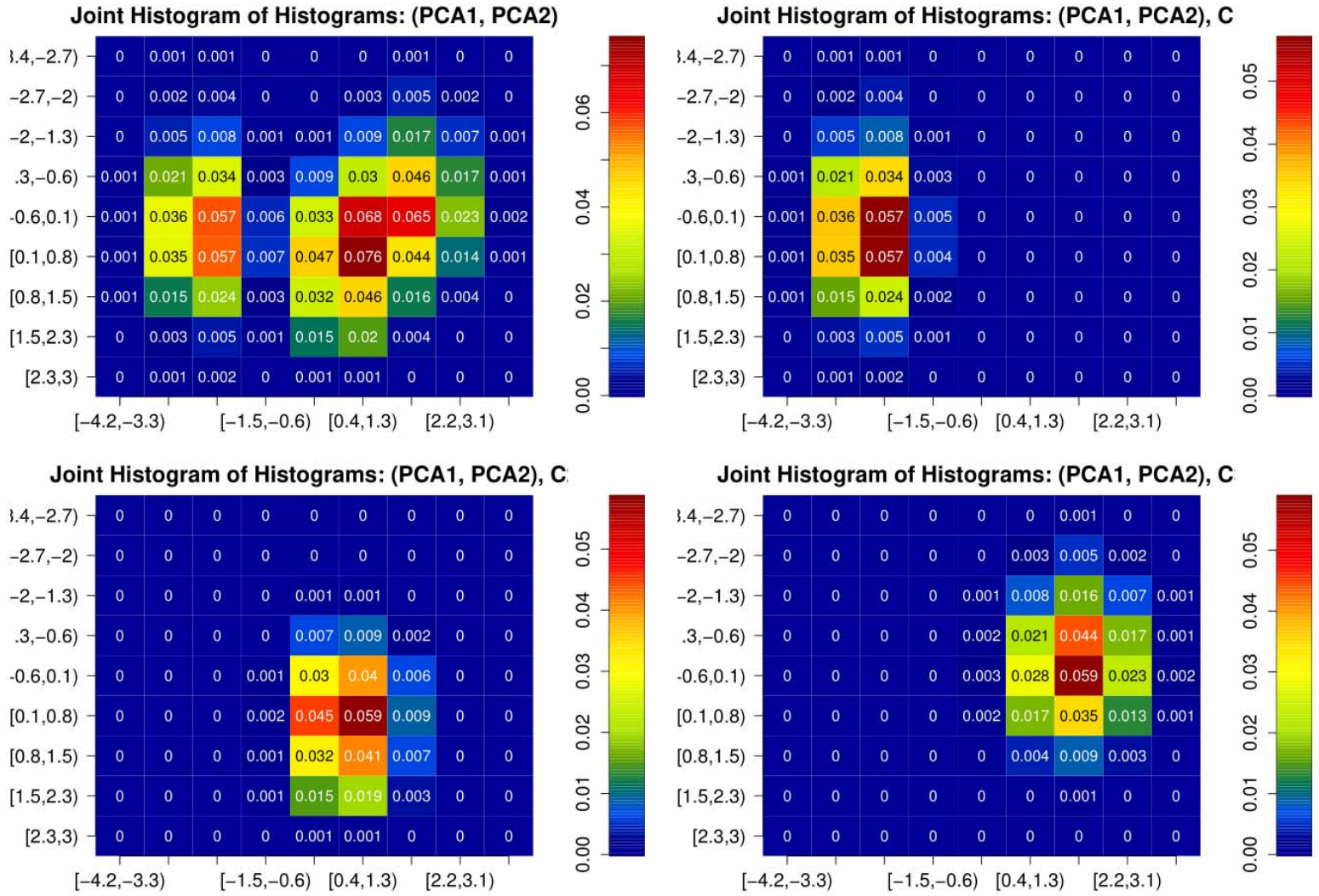
scatterplot3d.h(hdata.s2)



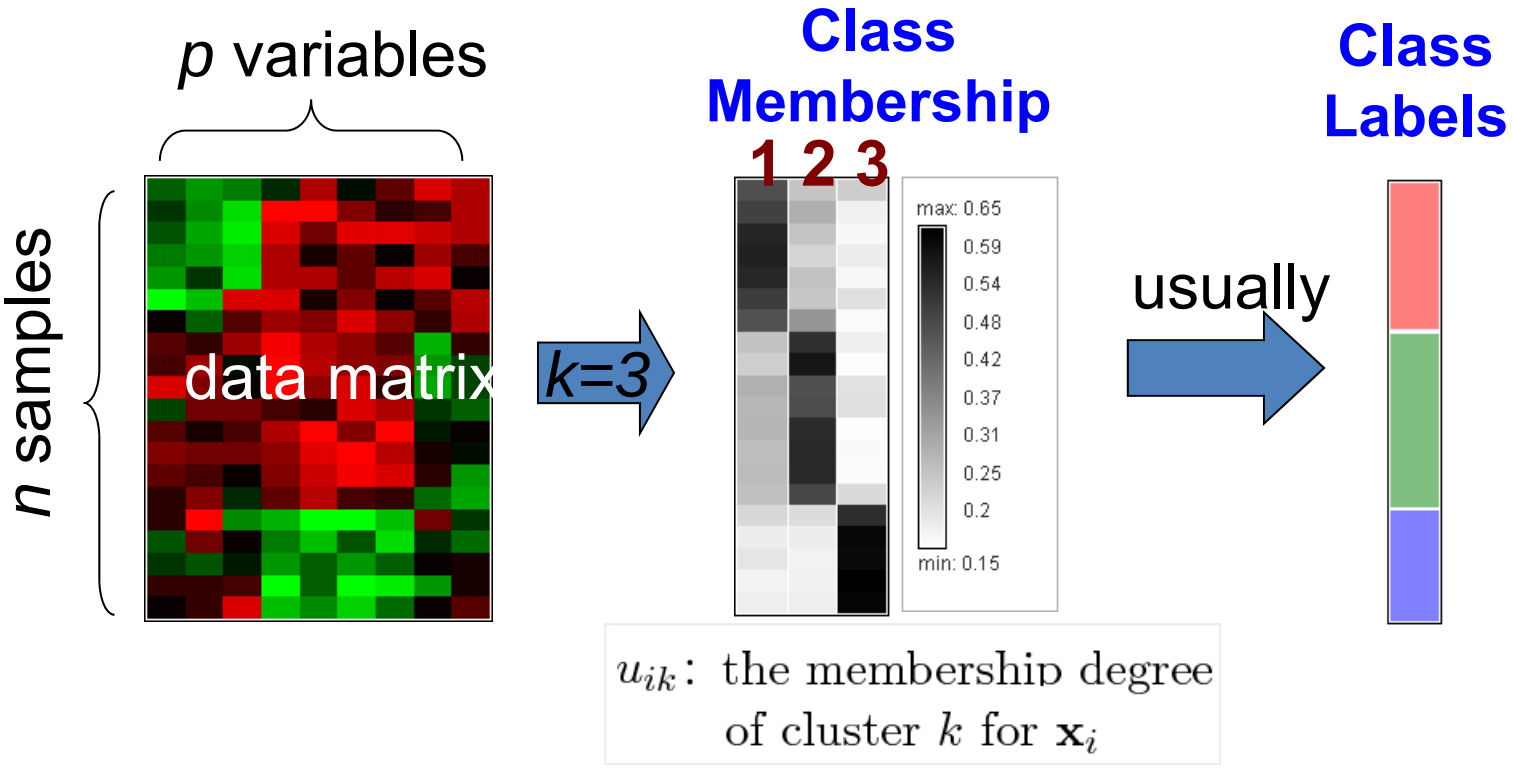




PCA for Histogram-valued Data

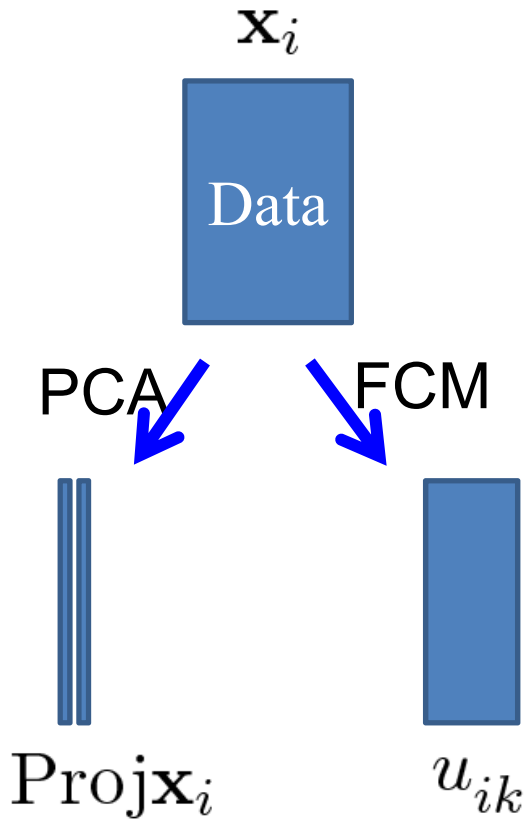


Visualization of Fuzzy (Soft) Clustering Results



- **Fuzzy Clustering methods:**
  Fuzzy c-means, model-based clustering, ensemble clustering,...
- **Visualization methods:** ?

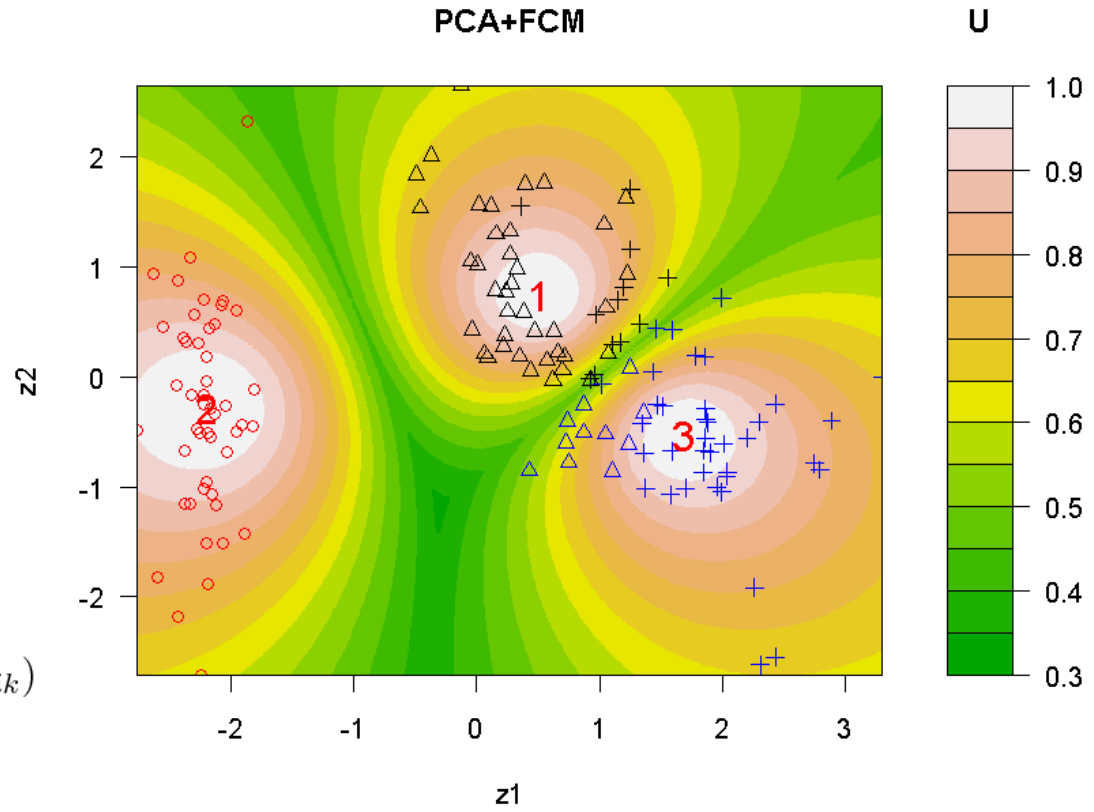
Visualization Method



$$\tilde{u}_{ik} = \left(\sum_{j=1}^{K} \left[\frac{d(\mathbf{g}_i, \mathrm{Proj}\mathbf{c}_k)}{d(\mathbf{g}_i, \mathrm{Proj}\mathbf{c}_j)} \right]^{\frac{2}{m-1}} \right)^{-1}$$

$$\mathrm{Proj}\mathbf{c}_k = \frac{\sum_{i=1}^{n} u_{ik}^m \cdot \mathrm{Proj}\mathbf{x}_i}{\sum_{i=1}^{n} u_{ik}^m}$$

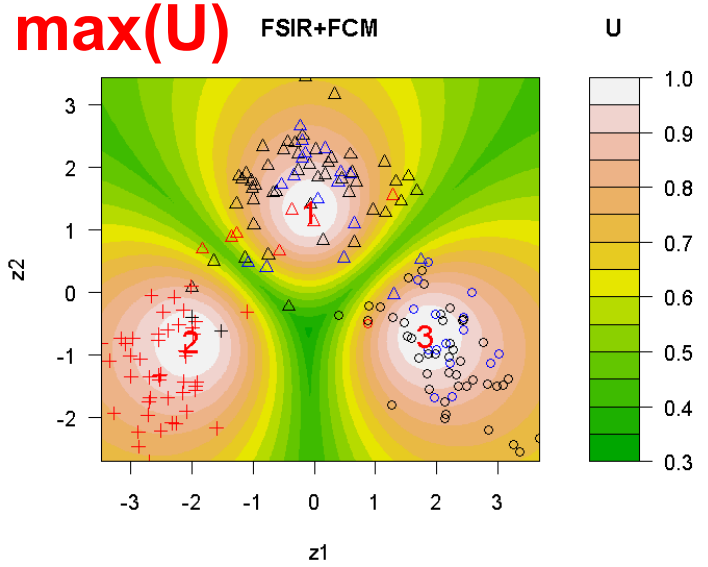
$$\tilde{y}_i = \arg\max_k (u_{ik})$$



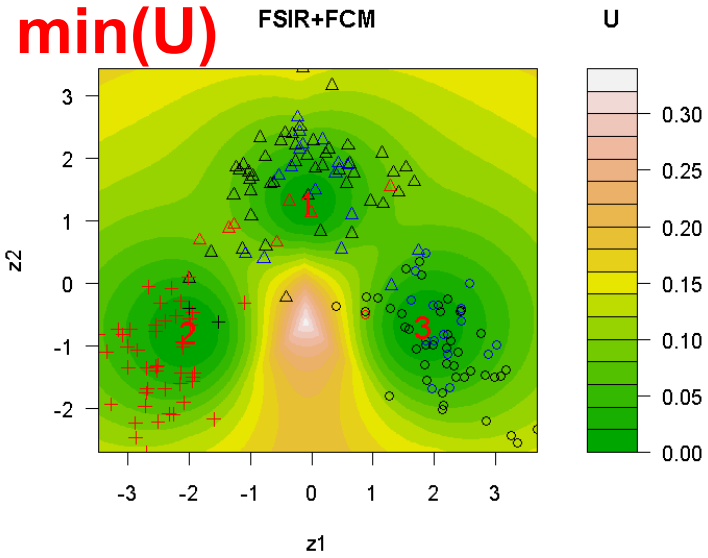
Example: Wine Data

(178x13: K=3)

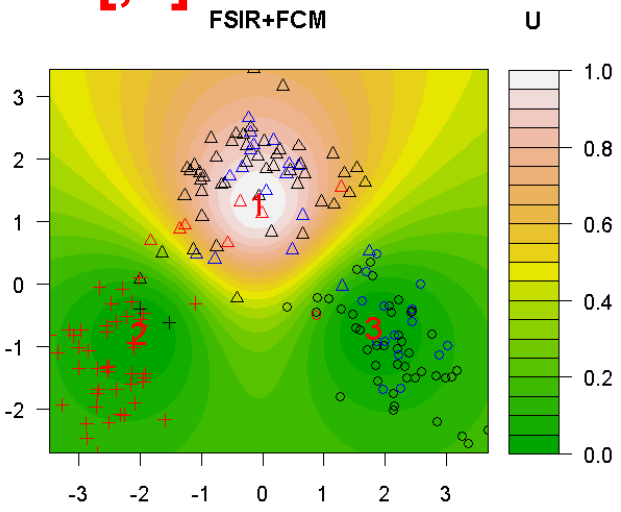
max(U)



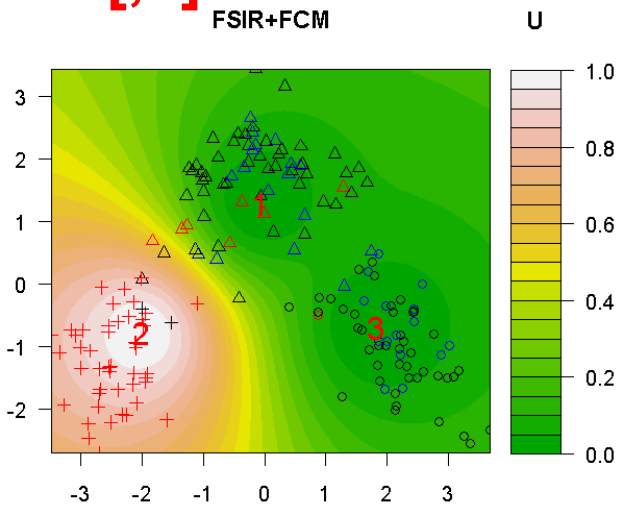
min(U)



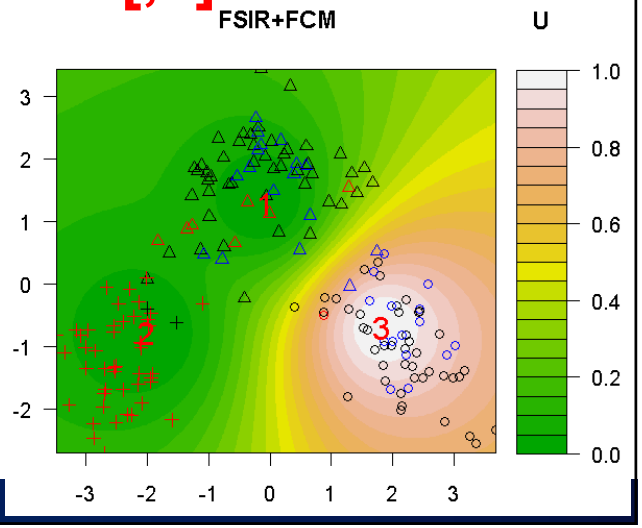
U[,1]



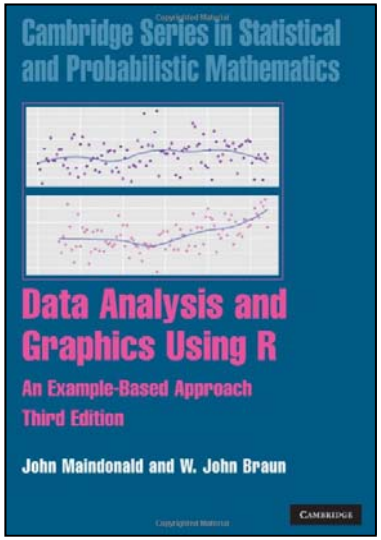
U[,2]

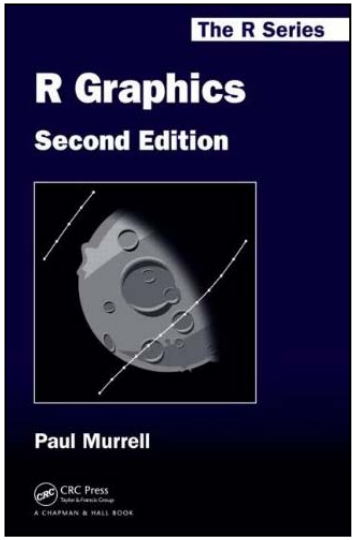


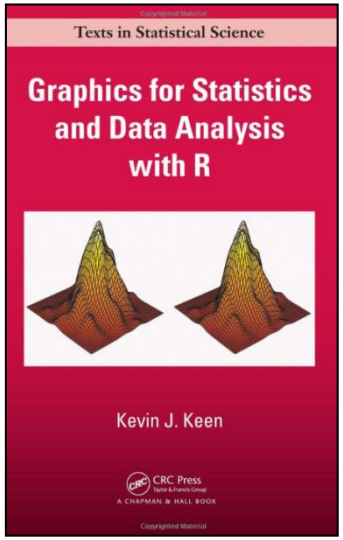
U[,3]

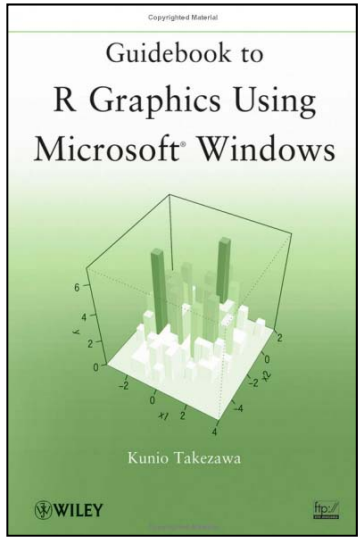


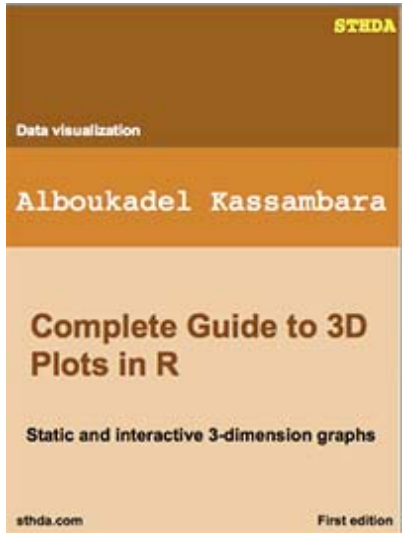
R圖形參考書目

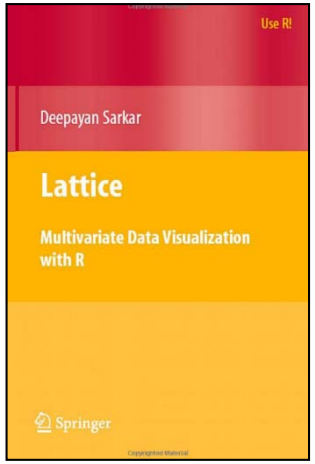

Cambridge Series in Statistical and Probabilistic Mathematics
Data Analysis and Graphics Using R
An Example-Based Approach
Third Edition
John Maindonald and W. John Braun
CAMBRIDGE

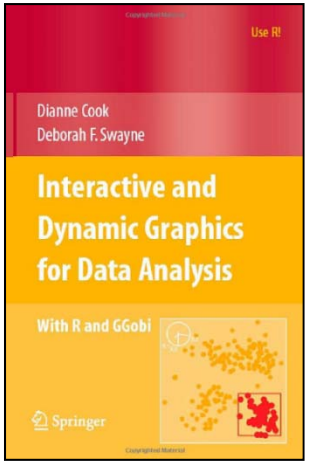

The R Series
R Graphics
Second Edition
Paul Murrell
CRC Press
A CHAPMAN & HALL BOOK

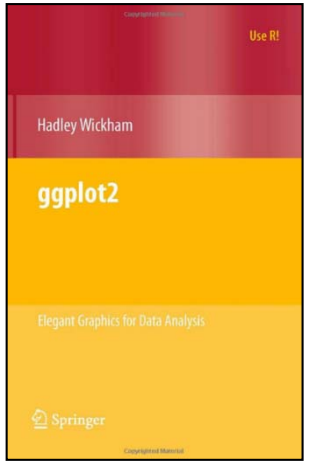

Texts in Statistical Science
Graphics for Statistics and Data Analysis with R
Kevin J. Keen
CRC Press
A CHAPMAN & HALL BOOK

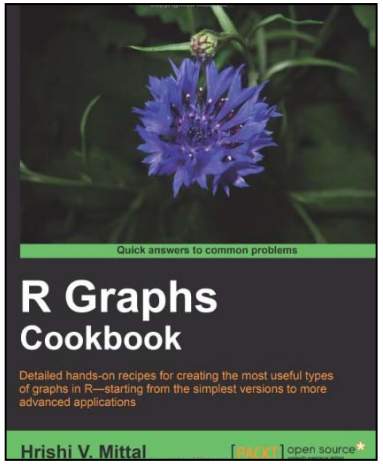

Guidebook to
R Graphics Using
Microsoft® Windows
Kunio Takezawa
WILEY

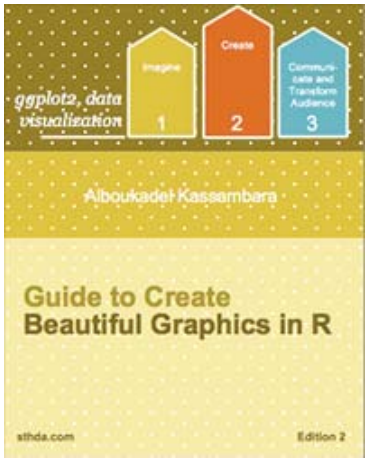

STHDA
Data visualization
Alboukadel Kassambara
Complete Guide to 3D Plots in R
Static and interactive 3-dimension graphs
sthda.com
First edition


Use R!
Deepayan Sarkar
Lattice
Multivariate Data Visualization with R
Springer

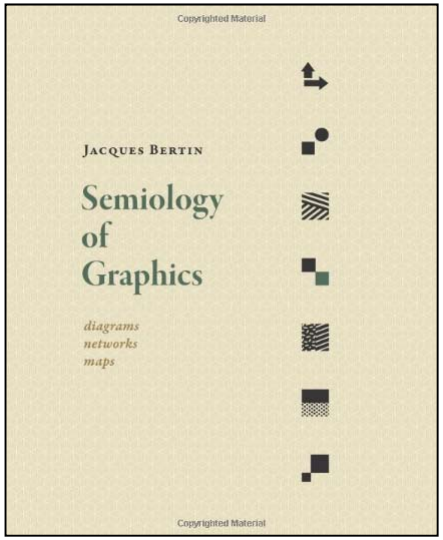

Use R!
Dianne Cook
Deborah F. Swayne
Interactive and Dynamic Graphics for Data Analysis
With R and GGobi
Springer

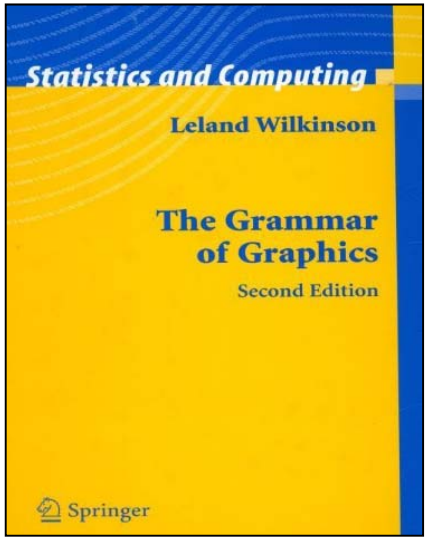

Use R!
Hadley Wickham
ggplot2
Elegant Graphics for Data Analysis
Springer

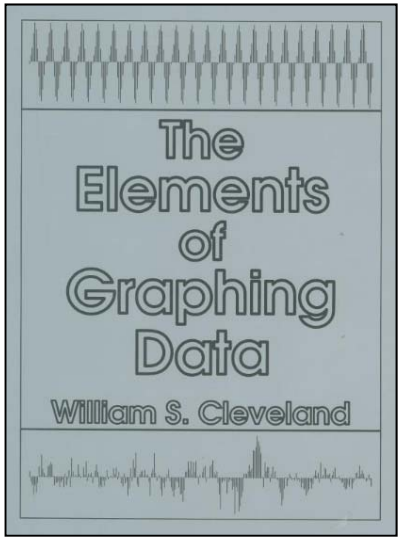

Quick answers to common problems
R Graphs
Cookbook
Detailed hands-on recipes for creating the most useful types of graphs in R—starting from the simplest versions to more advanced applications
Hrishi V. Mittal
open source

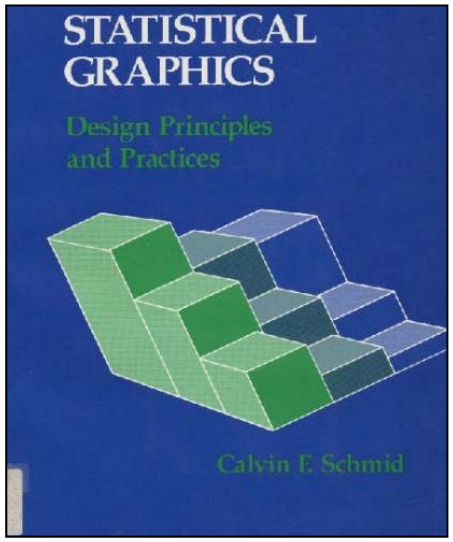

ggplot2, data visualization
Imagine
Create
Communicate and Transform Audience
1
2
3
Alboukadel Kassambara
Guide to Create Beautiful Graphics in R
sthda.com
Edition 2

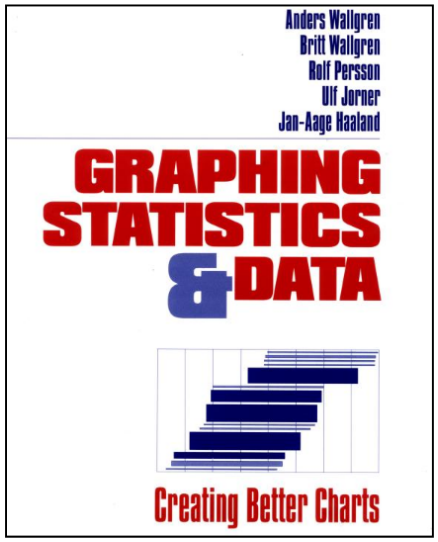
Fundations, Principle



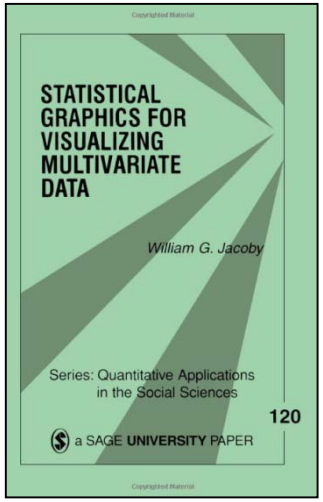


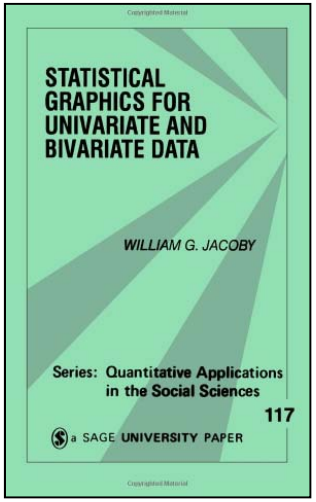


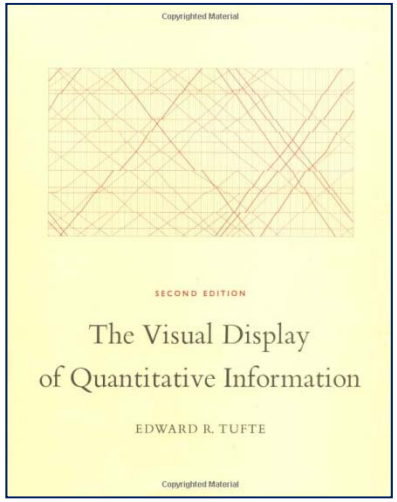


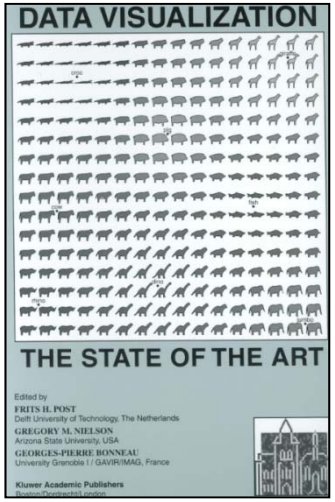


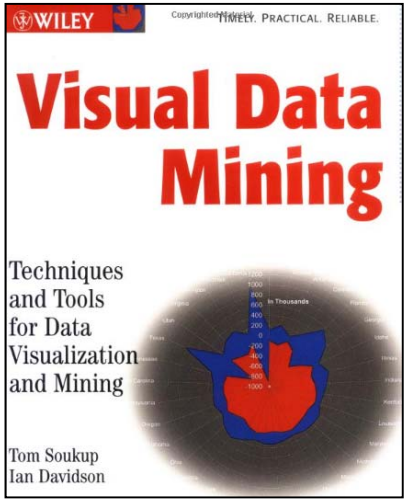
Statistics and Graphics

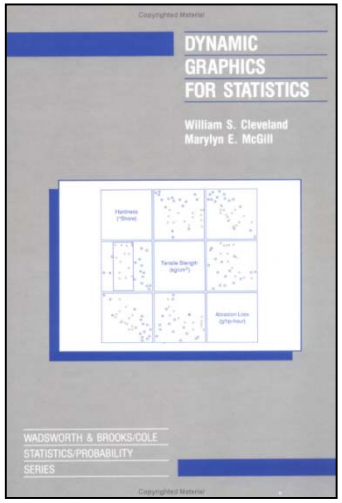


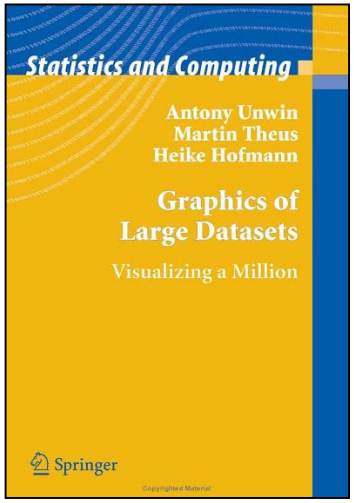


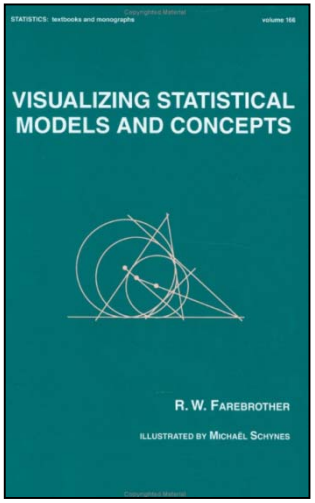


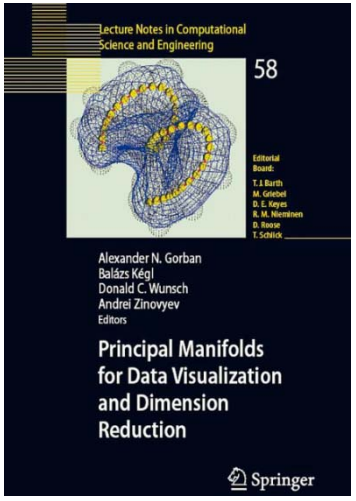


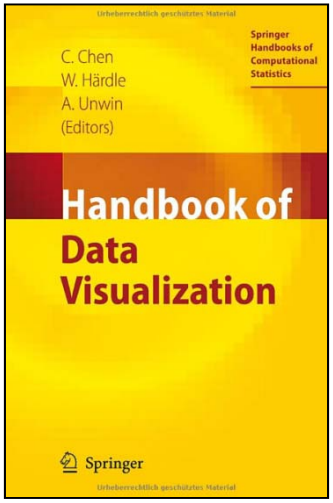




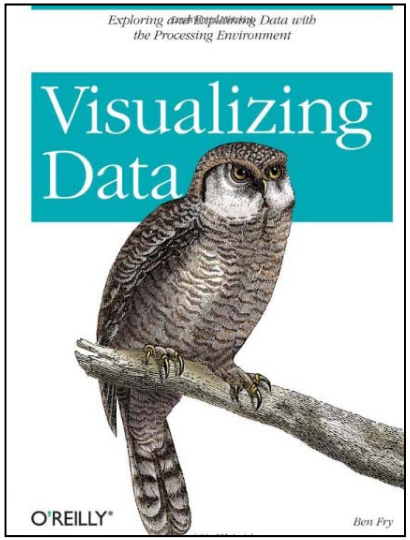


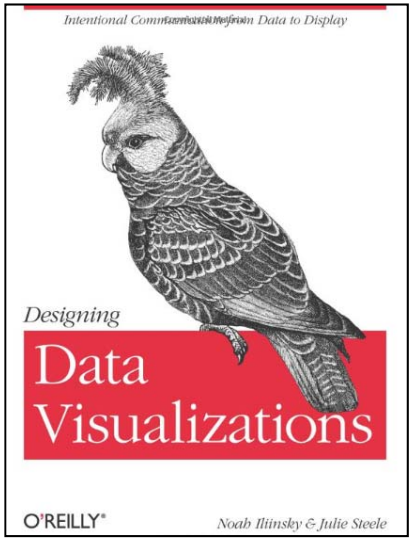


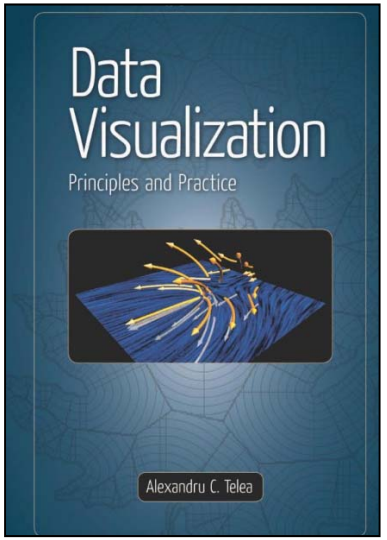


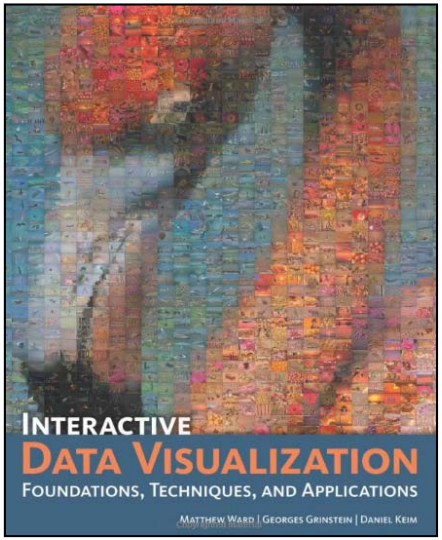


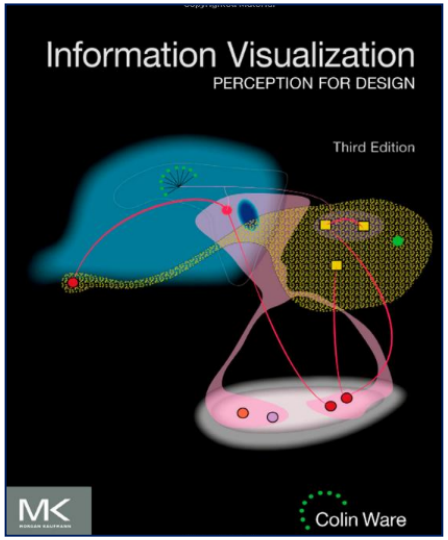
Data Visualization

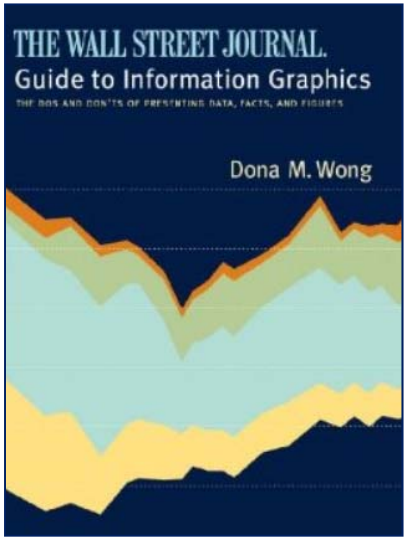


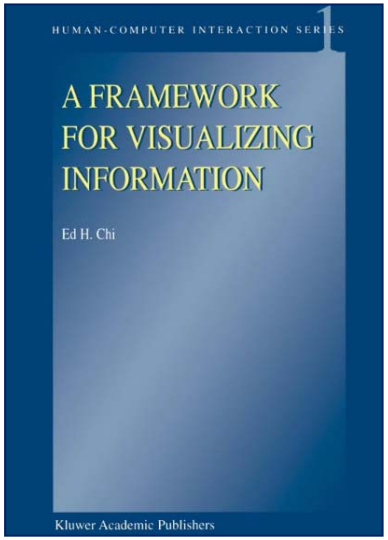




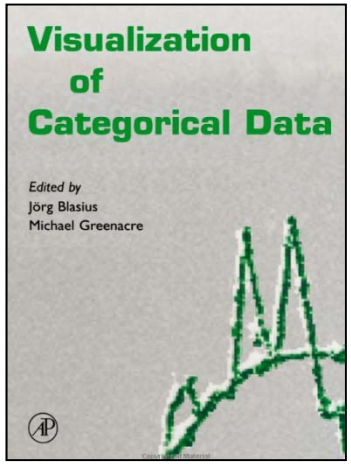


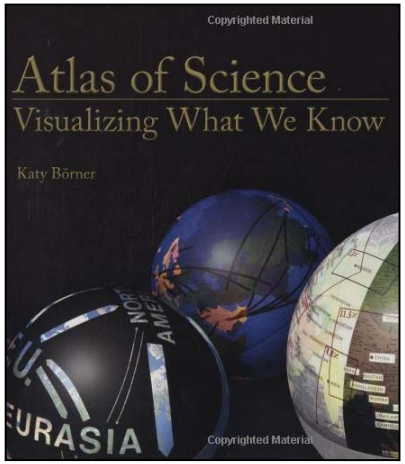


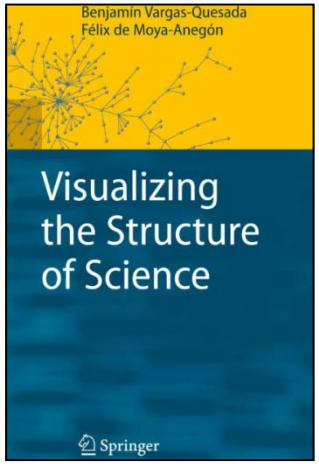


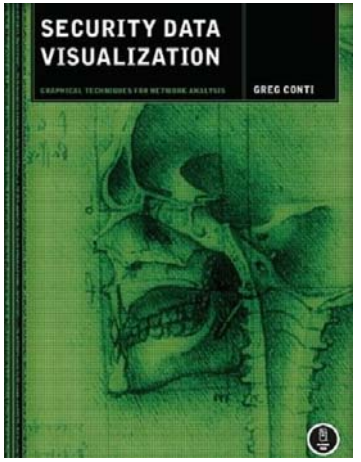


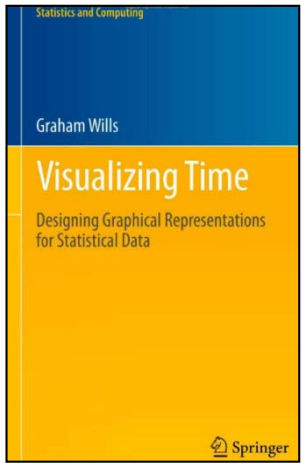
Specified Data Types

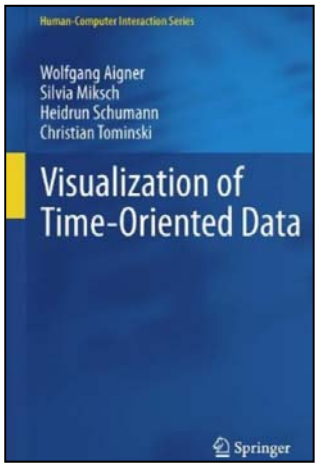


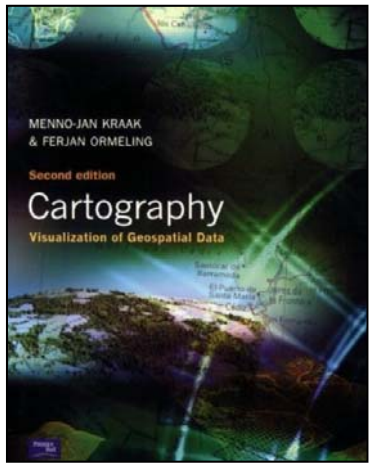


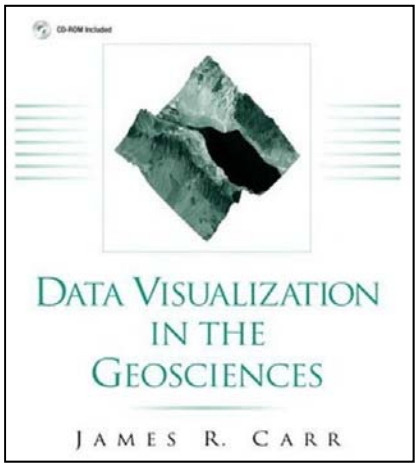


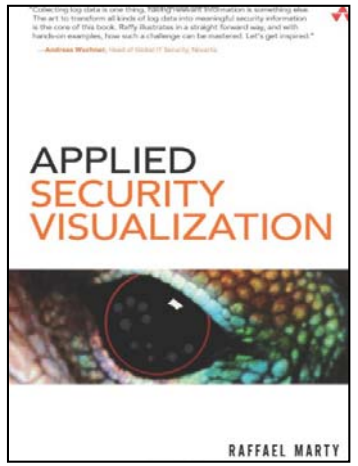




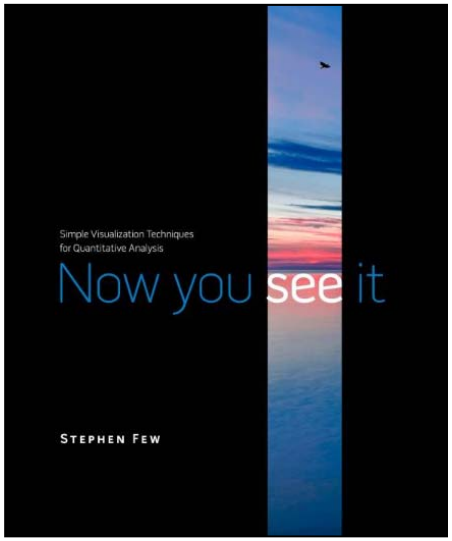


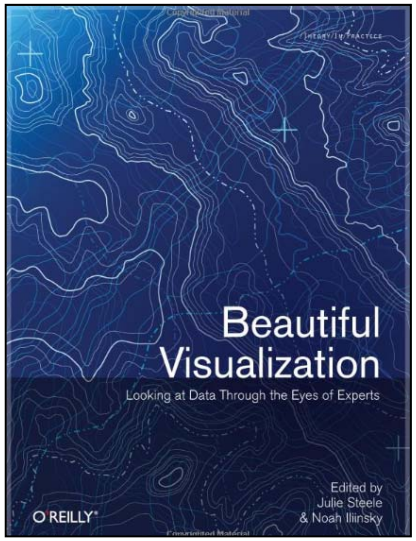


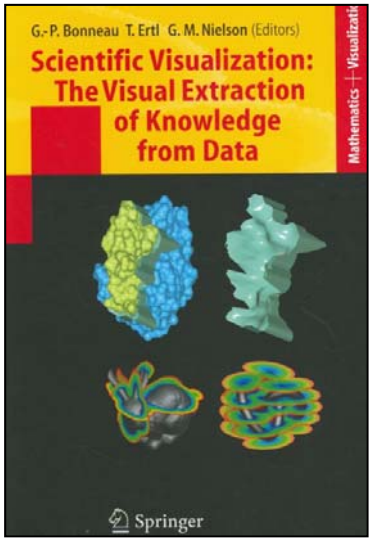


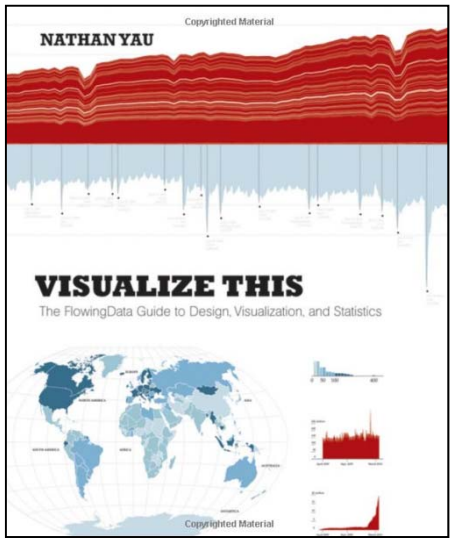


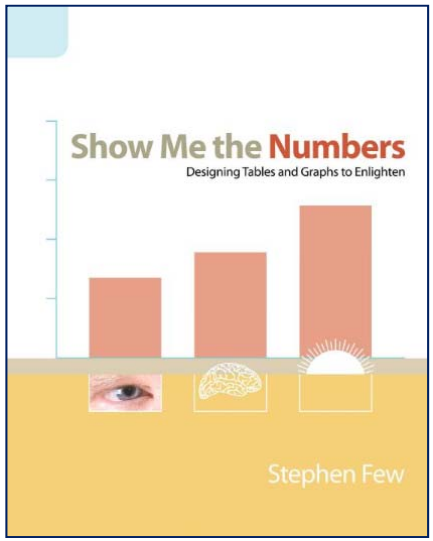
Others

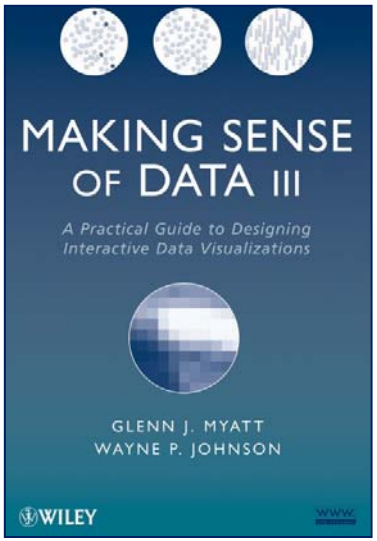

Simple Visualization Techniques
for Quantitative Analysis
Now you see it
Stephen Few


Beautiful
Visualization
Looking at Data Through the Eyes of Experts
Edited by
Julie Steele
& Noah Iliinsky
O'REILLY


G.-P. Bonneau T. Ertl G. M. Nielson (Editors)
Scientific Visualization:
The Visual Extraction
of Knowledge
from Data
Springer

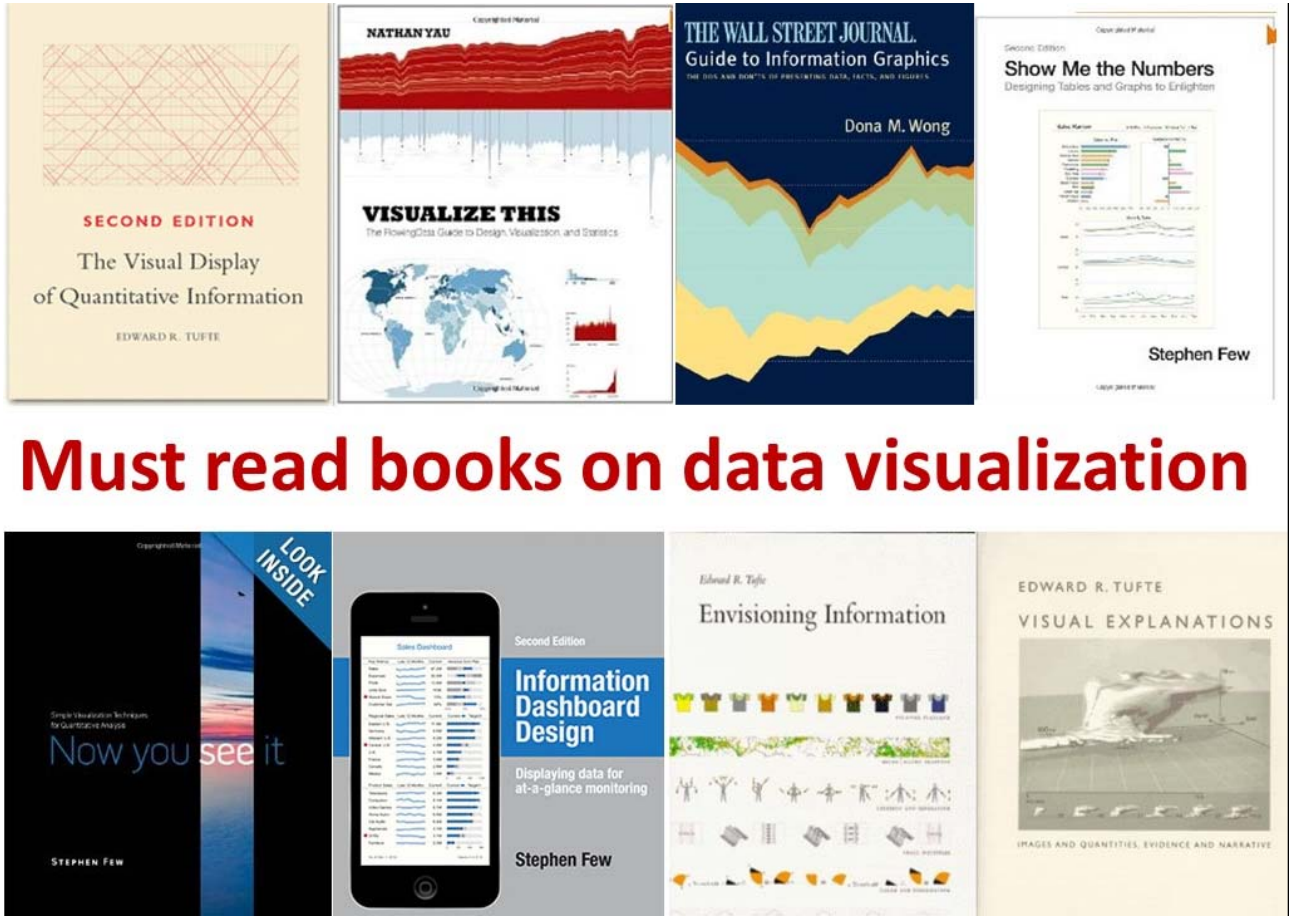

Copyrighted Material
NATHAN YAU
VISUALIZE THIS
The FlowingData Guide to Design, Visualization, and Statistics
Copyrighted Material


Show Me the Numbers
Designing Tables and Graphs to Enlighten
Stephen Few


MAKING SENSE
OF DATA III
A Practical Guide to Designing
Interactive Data Visualizations
GLENN J. MYATT
WAYNE P. JOHNSON
WILEY

Must read books on data visualization

https://www.analyticsvidhya.com/blog/2013/09/read-books-visualization/

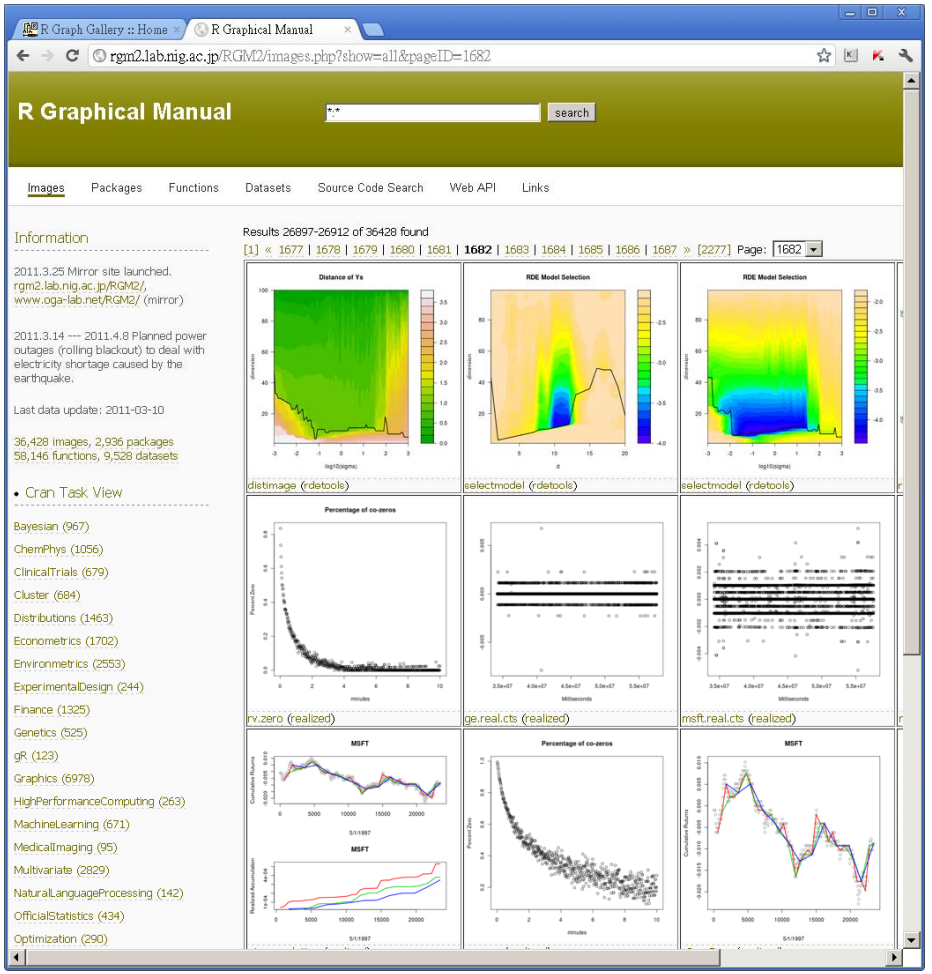


Sadly, not many analysts spend time on improving and thinking about visualization.

- Data Visualization – How to Pick the Right Chart Type?
https://eazybi.com/blog/data_visualization_and_chart_types/
- Introduction to Data Visualization: Chart Dos and Don'ts
http://guides.library.duke.edu/datavis/topten

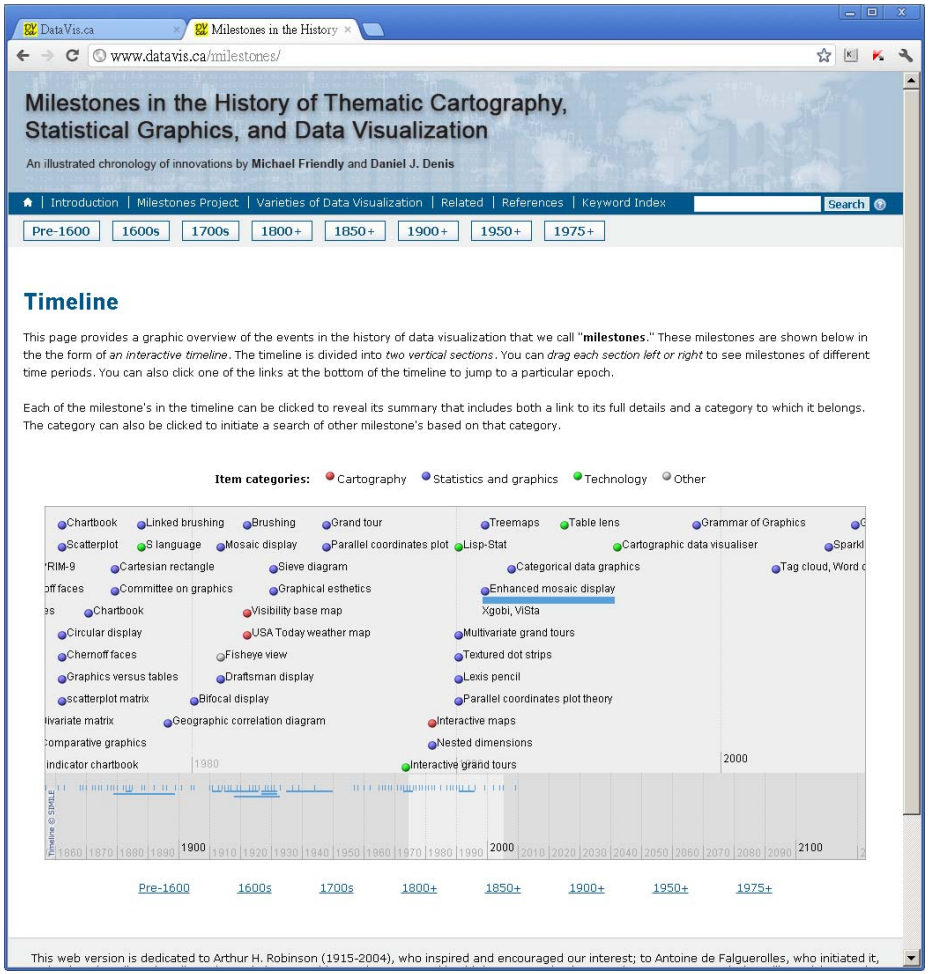
R Graphics


R Graph Gallery :: Home
R Graphical Manual
addictedtor.free.fr/graphiques/
RELATED SITES: R-project | CRAN | Bioconductor | R Wiki
Graph Gallery : Enhance your data visualization with R
search :
Donate
Home
List View
Thumbnails View
» Last entries ...
» Random entries
R Project
R is a system for statistical computation and graphics. It consists of a language plus a run-time environment with graphics, a debugger, access to certain system functions, and the ability to run programs stored in script files.
R Graphic Engine
One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.
Tweet
Like
Send
Tweets about the R Graph Gallery
Join the conversation
Recent news from R-bloggers
Quick View on Correlations of Different Instruments
http://www.r-bloggers.com/
Find us on Facebook
R Graph Gallery
Like
353 people like R Graph Gallery.
Facebook social plugin

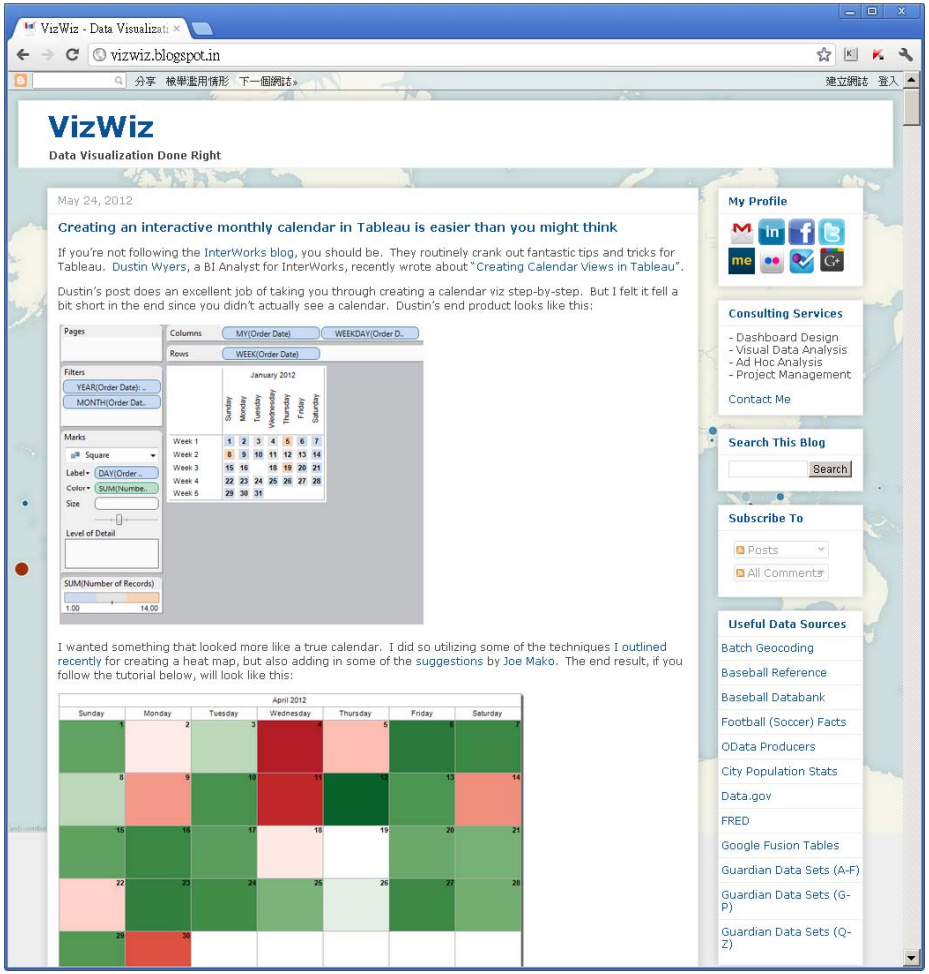

R Graph Gallery :: Home
R Graphical Manual
rgm2.lab.nig.ac.jp/RGM2/images.php?show=all&pageID=1682
R Graphical Manual
search
Images
Packages
Functions
Datasets
Source Code Search
Web API
Links
Information
2011.3.25 Mirror site launched.
rgm2.lab.nig.ac.jp/RGM2/,
www.oga-lab.net/RGM2/ (mirror)
2011.3.14 --- 2011.4.8 Planned power outages (rolling blackout) to deal with electricity shortage caused by the earthquake.
Last data update: 2011-03-10
36,428 images, 2,936 packages
58,146 functions, 9,528 datasets
Cran Task View
Bayesian (967)
ChemPhys (1056)
ClinicalTrials (679)
Cluster (684)
Distributions (1463)
Econometrics (1702)
Environmetrics (2553)
ExperimentalDesign (244)
Finance (1325)
Genetics (525)
gR (123)
Graphics (6978)
HighPerformanceComputing (263)
MachineLearning (671)
MedicalImaging (95)
Multivariate (2829)
NaturalLanguageProcessing (142)
OfficialStatistics (434)
Optimization (290)
Results 26897-26912 of 36428 found
[1] « 1677 | 1678 | 1679 | 1680 | 1681 | 1682 | 1683 | 1684 | 1685 | 1686 | 1687 » [2277] Page: 1682
distimage (rdetools)
selectmodel (rdetools)
selectmodel (rdetools)
rv.zero (realized)
ge.real.cts (realized)
msft.real.cts (realized)

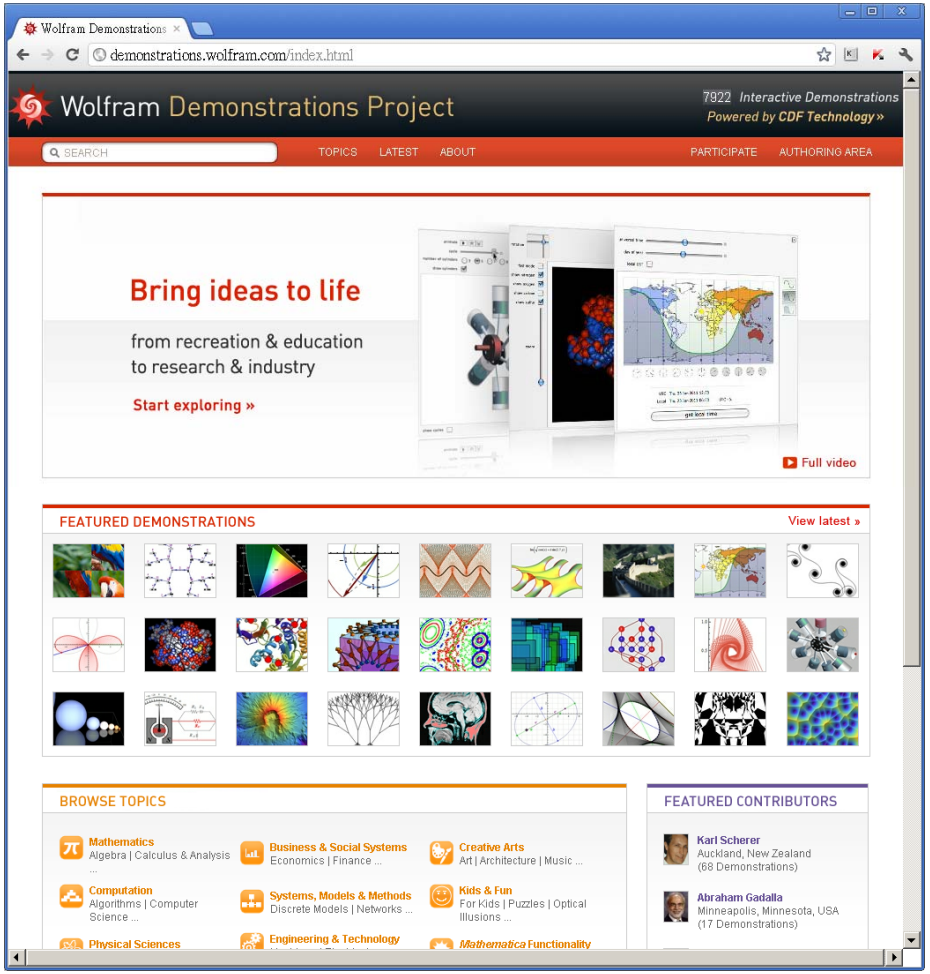
Web Resource (1)

Michael Friendly's Home Page

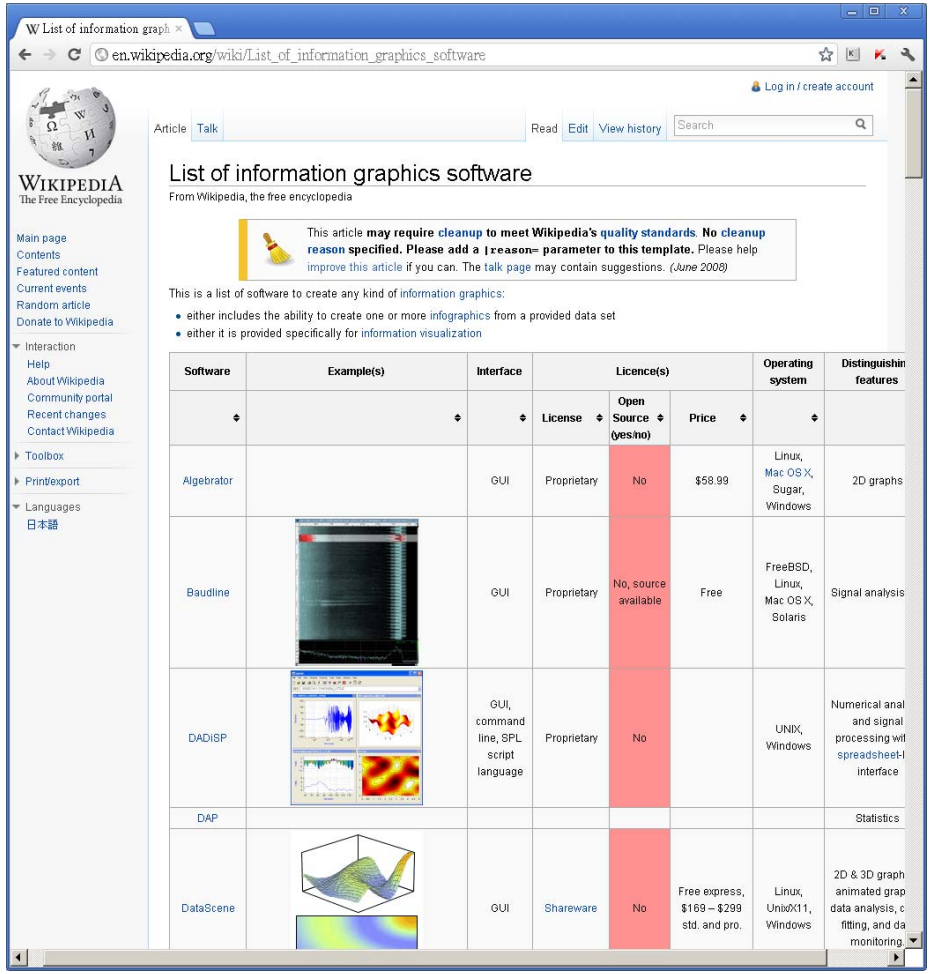


Web Resource (2)





List of Information Graphics Software



W List of information graph

en.wikipedia.org/wiki/List_of_information_graphics_software

Log in / create account

Article Talk Read Edit View history Search

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Toolbox
Print/export
Languages
日本語

List of information graphics software

From Wikipedia, the free encyclopedia

This article **may require cleanup to meet Wikipedia's quality standards. No cleanup reason specified. Please add a |reason= parameter to this template.** Please help improve this article if you can. The talk page may contain suggestions. *(June 2008)*

This is a list of software to create any kind of information graphics:

- either includes the ability to create one or more infographics from a provided data set
- either it is provided specifically for information visualization

Software	Example(s)	Interface	Licence(s)			Operating system	Distinguishin features
			License	Open Source (yes/no)	Price		
Algebrator		GUI	Proprietary	No	$58.99	Linux, Mac OS X, Sugar, Windows	2D graphs
Baudline		GUI	Proprietary	No, source available	Free	FreeBSD, Linux, Mac OS X, Solaris	Signal analysis
DADiSP		GUI, command line, SPL script language	Proprietary	No		UNIX, Windows	Numerical anal and signal processing wit spreadsheet-l interface
DAP							Statistics
DataScene		GUI	Shareware	No	Free express, $169 – $299 std. and pro.	Linux, UnixX11, Windows	2D & 3D graph animated grap data analysis, c fitting, and da monitoring.

Top 10 Graphical User In

www.decisionstats.com/top-10-graphical-user-interfaces-in-statistical-softwares/

Predictive Analytics WORLD
Strengthen the business impact delivered by analytics
LEARN MORE!
JUNE 25-26, 2012 | CHICAGO
INTERVIEWS

ABOUT DECISIONSTATS

SPONSOR: ANALYTICS TRAINING

JIGSAW ACADEMY

www.jigsawacademy.in

Online Analytics Training

Learn advanced analytics with tools such as SAS, R, Excel Angoss and SPSS

TOP 10 GRAPHICAL USER INTERFACES IN STATISTICAL SOFTWARE

APRIL 29, 2010 BY AJAY OHRI 9 COMMENTS

Flattr 0

Here is a list of top 10 GUIs in Statistical Software. The overall criterion is based on-

- User Friendly Nature for a New User to begin click and point and learn.
- Cleanliness of Automated Code or Log generated.
- Practical application in consulting and corporate world.
- **Cost** and **Ease** of Ownership (including purchase,install,training,maintainability,renewal)
- Aesthetics (or just plain pretty)

However this list is not in order of ranking- (as beauty (of GUI) lies in eyes of the beholder). For a list of top 10 GUI in R language only please see -

https://rforanalytics.wordpress.com/graphical-user-interfaces-for-r/

This is only a GUI based list so **it excludes** notable *command line* or *text editor submit commands* based softwares which are also very powerful and user friendly.

1. **JMP** -

While critics of SAS Institute often complain on the premium pricing of the basic model (especially AFTER the entry of another SAS language software WPS from http://www.teamwpc.co.uk/products/wps – they should try out JMP from http://jmp.com – it has a 1 month free evaluation, is much less expensive and the GUI makes it very very easy to do basic statistical analysis and testing. The learning curve is surprisingly fast to pick it up (as it should be for well designed interfaces) and it allows for very good quality output graphics as well.

CONFERENCES

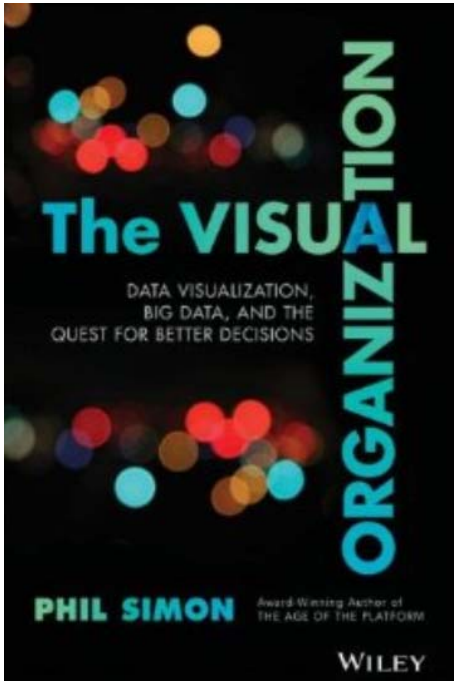
Predictive Analytics WORLD
JUNE 25-26
CHICAGO
Strengthen the business impact delivered by analytics
LEARN MORE!

Some DataViz Sites

- Data Visualization Is The Future - Here's Why

http://www.forbes.com/sites/dorieclark/2014/03/10/data-visualization-is-the-future-heres-why/

- Phil Simon, 2014, The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions, Wiley.
  ISBN: 9781118794388 | 1118794389



- Information Aesthetics: http://infosthetics.com/
- Chart Porn: http://chartporn.org/
- Eagereyes: https://eagereyes.org/
- We Love Datavis: http://datavis.tumblr.com
- A New Generation Tool For (big) Data Visualization: http://www.stratio.com/datavis/kbase/
- Visualizing.org: http://www.visualizing.org/explore
- VizWiz: http://vizwiz.blogspot.ca/
- US Census Data Visualization Gallery: http://www.census.gov/dataviz/

Acknowledgment

TKU, NTPU, MOST, Chun-houh Chen (中研院統計所 陳君厚)

Han-Ming Wu 吳漢銘

http://www.hmwu.idv.tw

國立臺北大學 統計學系

Department of Statistics, National Taipei University

Home | About Me | Photo Gallery | Facebook | Links | Contact Me | NTPU-106(上)課程 | 【作業考試上傳區】

NTPU-106(上)課程

- ✓ 微積分
- ✓ 高維度資料分析
- 【作業考試上傳區】
- 【歷年課程】

NTPU-106(下)課程

- ✓ 微積分
- ✓ 電腦概論與程式設計

460146

Today	32
This Week	3602
This Month	14298
All days	460146

Your IP: 1.34.216.123

2018-01-28 02:46

Visitors Counter

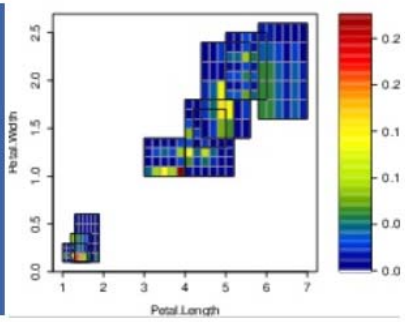
Home

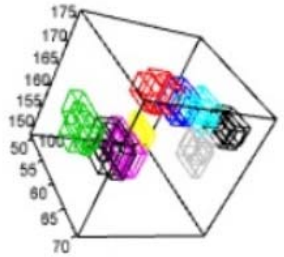
iasc isi International Association for Statistical Computing

IASC-ARS The Asian Regional Section of the International Association for Statistical Computing

iasc The Young Statisticians Group in The International Association for Statistical Computing

YS The Young Statisticians International Statistical Institute http://www.ys-isi.org





R Software (R統計軟體教學)

Exploratory Symbolic Data Analysis

Teaching 教學

- Courses (歷年課程)
- R Software (R統計軟體教學)

Research 研究

- Publication (發表)
- Research (研究)

Service 服務

- Journal Referee (學術審稿)
- Alumni.Math (數學系系友會)