Visualization & EDA in R

An introduction to the R statistical framework

Data Visualization

Prof. Dr. Javier Valdes
Javier.valdes@th-deg.de
Technische Hochschule Deggendorf

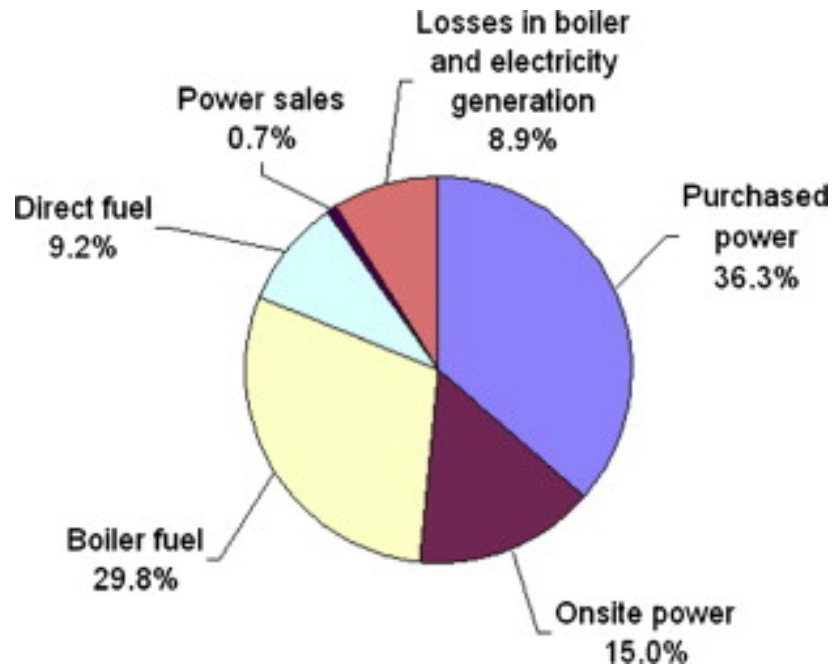# Energy flow analysis in pulp and paper industry

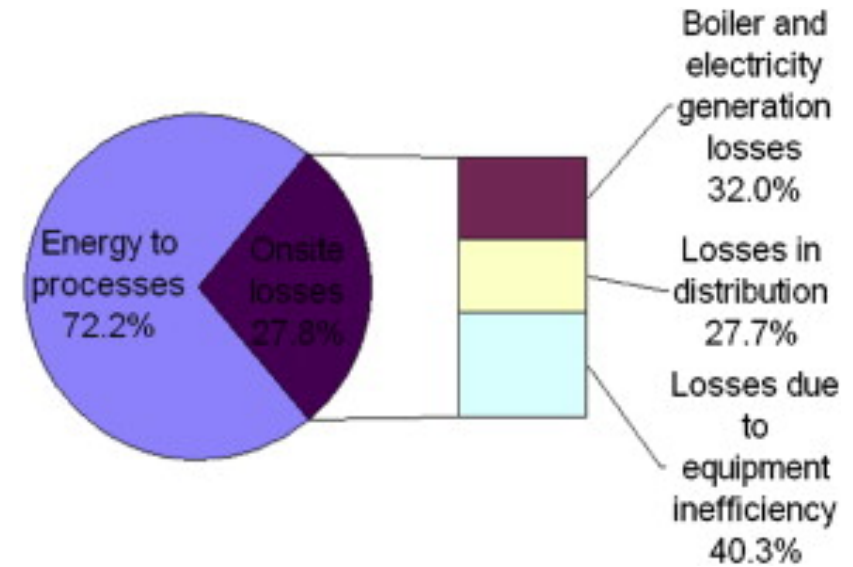Fig. 3. Primary energy use distribution of Taiwanese pulp and paper industry

Fig. 4. Onsite energy loss profile of Taiwanese pulp and paper industry.

Hong, Gui-Bing, et al. "Energy flow analysis in pulp and paper industry." *Energy* 36.5 (2011): 3063-3068.

An extensive comparative study of cluster validity indices

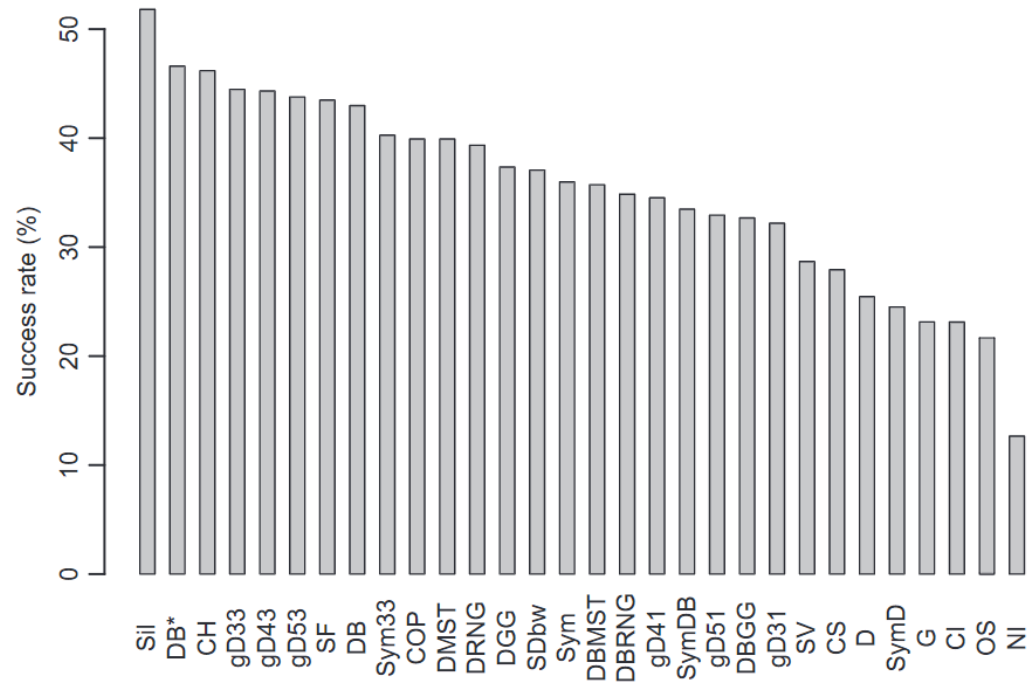O. Arbelaitz et al. / Pattern Recognition 46 (2013) 243–256

Fig. 2. Overall results for the experiment with synthetic datasets.

Arbelaitz, Olatz, et al. "An extensive comparative study of cluster validity indices." *Pattern Recognition* 46.1 (2013): 243-256.

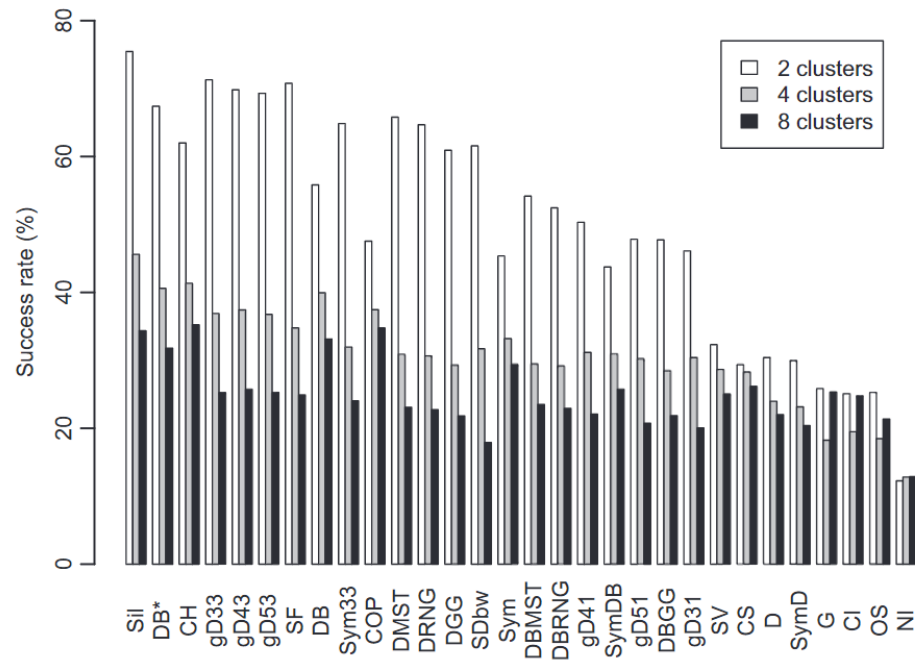# An extensive comparative study of cluster validity indices

Fig. 4. Results for synthetic datasets broken down by number of clusters.
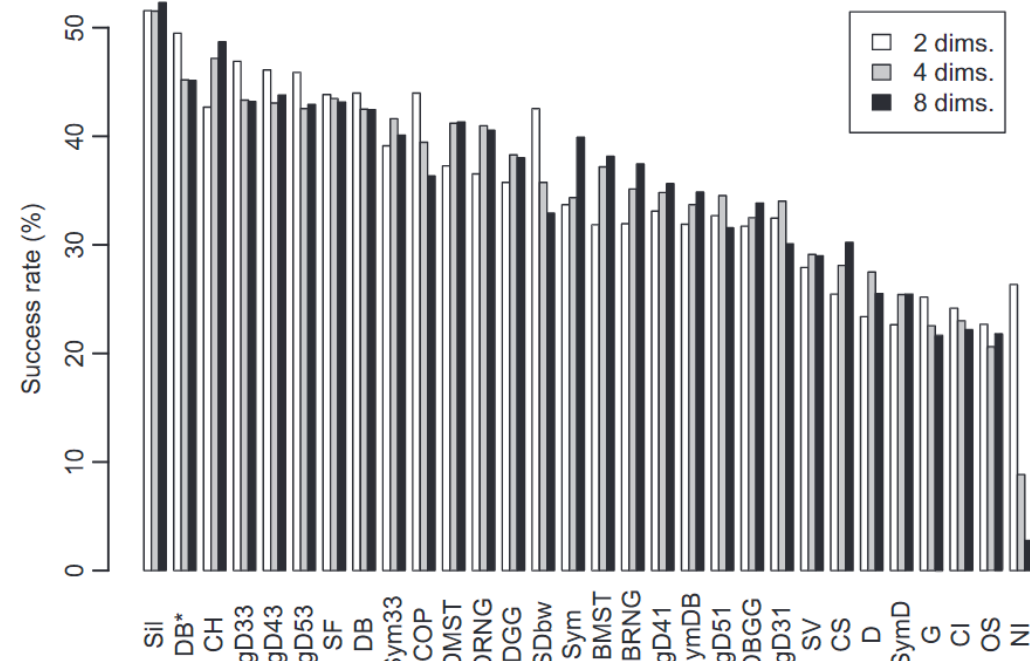

Fig. 5. Results for synthetic datasets broken down by dimensionality.

Arbelaitz, Olatz, et al. "An extensive comparative study of cluster validity indices." *Pattern Recognition* 46.1 (2013): 243-256.

Plotting

Base plot types

Function Graph type
plot()   Scatter plots and line plots
barplot() Bar plot, with a unending number of options
hist()   histograms and relative frequency diagrams
curve()   plots curves given some mathematical expression
boxplot() box-and-wisker plots
symbols() as scatter plots but symbols sized by other variable

```r
# initialization

par(mar=c(3,3,3,3))
num <- 0 ;
num1 <- 0
plot(0,0 , xlim=c(0,21) , ylim=c(0.5,6.5), col="white" , yaxt="n" , ylab="" , xlab="")

#fill the graph
for (i in seq(1,20)){
    points(i,1 , pch=i , cex=3)
    points(i,2 , col=i , pch=16 , cex=3)
    points(i,3 , col="black" , pch=16 , cex=i*0.25)

#lty

if(i %in% c(seq(1,18,3))){
    num= num+1
    points(c(i,i+2), c(4,4) , col="black" , lty=num , type="l" , lwd=2)
    text(i+1.1 , 4.15 , num)
    }

#type and lwd
if(i %in% c(seq(1,20,5))){
    num1=num1+1
    points(c(i,i+1,i+2,i+3), c(5,5,5,5), col="black", type= c("p","l","b","o") [num1] , lwd=2)
    text(i+1.1 , 5.2 , c("p","l","b","o")[num1] )
    points(c(i,i+1,i+2,i+3), c(6,6,6,6) , col="black" , type="l", lwd=num1)
    text(i+1.1 , 6.2 , num1 ) } }

#add axis

axis(2, at = c(1,2,3,4,5,6), labels = c("pch" , "col" , "cex" , "lty", "type" , "lwd" ), tick =
TRUE, col = "black", las = 1, cex.axis = 0.8)
```

6

```r
# initialization

par(mar=c(3,3,3,3)) # multiple graphs in a single plot  # A numerical vector of the form c(bottom, left, top, right) which gives the number of lines of margin to be specified on the
four sides of the plot
num <- 0 ;
num1 <- 0
plot(0,0 , xlim=c(0,21) , ylim=c(0.5,6.5), col="white" , yaxt="n" , ylab="" , xlab="")

#fill the graph
for (i in seq(1,20)){
    points(i,1 , pch=i , cex=3)
    points(i,2 , col=i , pch=16 , cex=3)
    points(i,3 , col="black" , pch=16 , cex=i*0.25)

#lty

if(i %in% c(seq(1,18,3))){
    num= num+1
    points(c(i,i+2), c(4,4) , col="black" , lty=num , type="l" , lwd=2)
    text(i+1.1 , 4.15 , num)
    }

#type and lwd
if(i %in% c(seq(1,20,5))){
    num1=num1+1
    points(c(i,i+1,i+2,i+3), c(5,5,5,5), col="black", type= c("p","l","b","o") [num1] , lwd=2)
    text(i+1.1 , 5.2 , c("p","l","b","o")[num1] )
    points(c(i,i+1,i+2,i+3), c(6,6,6,6) , col="black" , type="l", lwd=num1)
    text(i+1.1 , 6.2 , num1 ) } }

#add axis

axis(2, at = c(1,2,3,4,5,6), labels = c("pch" , "col" , "cex" , "lty", "type" , "lwd" ), tick = TRUE, col = "black", las = 1, cex.axis = 0.8)
```
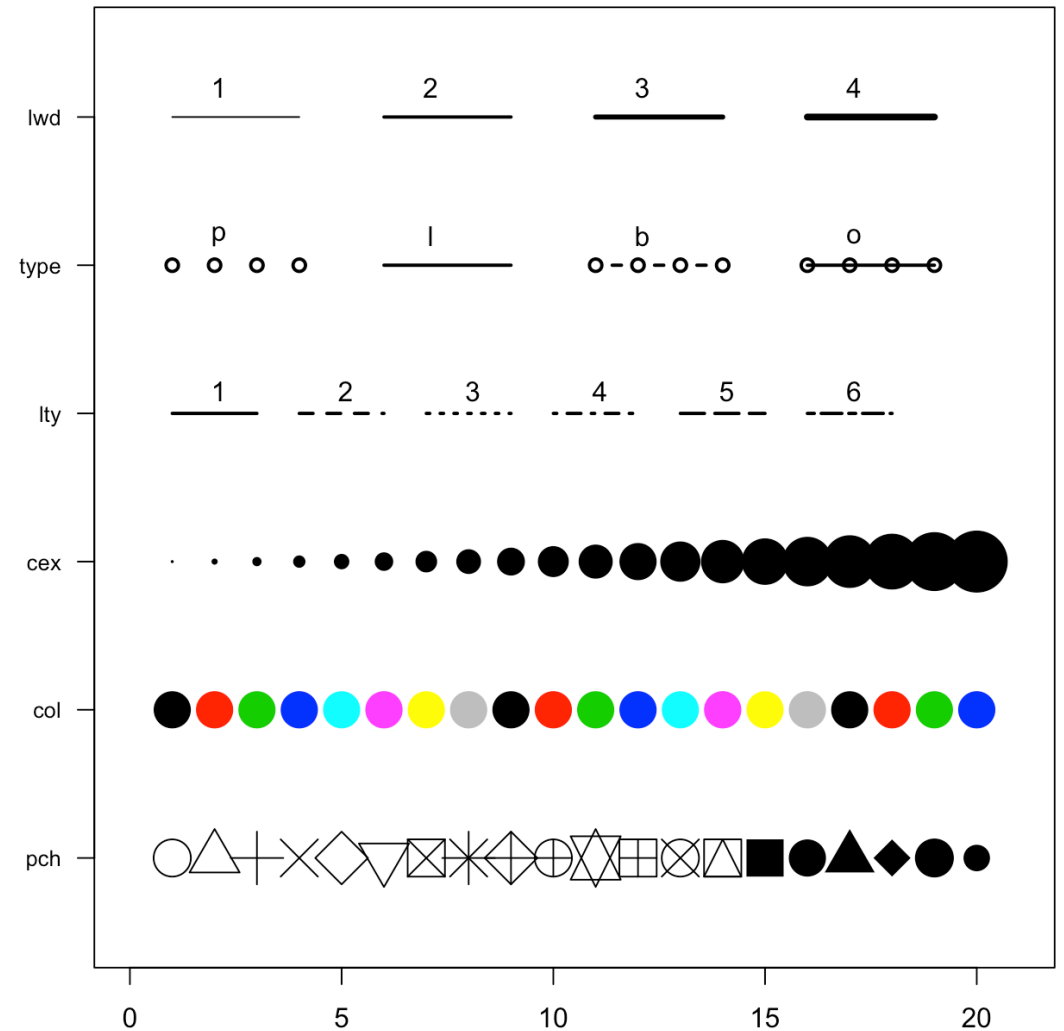
Plotting

SET GRAPHICAL PARAMETERS

*the following can **only** be set with **par()***

par (...)

multiple plots	`mfcol = c(`*nrow,ncol*`)` `mfrow = c(`*nrow,ncol*`)`	*plot margins (outer)*	`oma = c(`*bottom, left, top, right*) default: `c(0, 0, 0, 0)` lines
plot margins	`mar = c(`*bottom, left, top, right*) default: `c(5.1, 4.1, 4.1, 2.1)` lines	*query x & y limits*	`par ("usr")`

Plotting

Bar charts   **barplot**(*height, …*)
- *bar labels*   `names.arg =`
- *border*   `border =`
- *fill color*   `col =`
- *horizontal*   `horiz = TRUE`

Box plots   **boxplot**(*x, …*)
- *horizontal*   `horizontal = TRUE`
- *box labels*   `names =`

Dot plots   **dotchart**(*x, …*)
- *dot labels*   `labels =`

Histograms   **hist**(*x, …*)
- *breakpts*   `breaks =`

Line charts   **plot**(*x,* `type = "l"`)
- *line type*   `lty =`   "blank"|0
      "solid"|1
      "dashed"|2
      "dotted"|3
- *line width*   `lwd =`

Scatterplots   **plot**(*x, …*)
- *symbol*   `pch =`

Plotting

REMOVE		ADJUST	
axis labels	`ann = FALSE`	*allow plotting out of plot region*	`xpd = TRUE`
axis, tickmarks, and labels	`xaxt = "n"` `yaxt = "n"`	*aspect ratio*	`asp =`
plot box	`bty = "n"`	*axis limits*	`xlim =, ylim =`
		axis lines to match axis limits	`xaxs = "i"` , `yaxs = "i"` *(internal axis calculation)*

NOTE: Many of the parameters here can be also be set in `par()`. *See R help for more options.*

Plotting

ADD TEXT

location

axis labels	`xlab =, ylab =`
subtitle	`sub =`
title	`main =`

style

font face	`font = 1` *(plain)*
	`2` *(bold)* `3` *(italic)*
	`4` *(bold italic)*
font family	`family = "serif"`
	`"sans"` `"mono"`

size

(magnification factor)

all elements	`cex =`
axis labels	`cex.lab =`
subtitle	`cex.sub =`
tick mark labels	`cex.axis =`
title	`cex.main =`

position

text direction	`las = 1` (horizontal)
justification	`adj = 0 .5 1`
	(left, center, right)

Plotting

Add new plot   [*any plot function*]
               (... , add = TRUE)
   ex. barplot(x, add = TRUE)

Axes   axis (*side,...*)
  location   side = 1 2 3 4
             (bottom, left, top, right)
tick mark:
  labels   labels =
  location   at =
  remove   tick = FALSE
  rotate text   las = 1 (horizontal)

Axis labels   mtext (*text,...*)
  location   side = 1 2 3 4
          (bottom, left, top, right)
  lines to skip   line = *(from plot*
              region, default = 0)
  position   at = *x or y-coord*
          (depending on side)
  justification   adj = 0 .5 1
          (left, center, right)

Lines   lines (*x,...*)
  line style   lty =
  line width   lwd =
  color   col =

Points   points (*x,...*)
  symbol   pch =

□ ○ △ ＋ × ◇ ▽ ⊠ ✳ ◈ ⊕ ⊠ ⊞
0 1 2 3 4 5 6 7 8 9 10 11 12

⊠ ◪ ■ ● ▲ ◆ ● ● ○ □ ◇ △ ▽
13 14 15 16 17 18 19 20 21 22 23 24 25

  color   col =
  fill color   bg = (pch: 21-25 only)

Text   text (*x, y, text,...*)
  position   pos = 1 2 3 4
  (rel. to x,y)   (below, left, above, right)
              (default=center)

Title   title (*main,...*)
  axis labels   xlab =, ylab =
  subtitle   sub =
  title   main =

Visualization & EDA in R

An introduction to the R statistical framework

Data Visualization

Prof. Dr. Javier Valdes
Javier.valdes@th-deg.de
Technische Hochschule Deggendorf