R – Commandes utiles

Rnorm(length,mean,var);

Convertir cellules d'un dataframe de factor à character

<http://stackoverflow.com/questions/19204729/how-to-change-factor-labels-into-string-in-a-data-frame>

Bien lire un fichier csv: indiquer le type des colonnes à la lecture

<http://stackoverflow.com/questions/2805357/specifying-colclasses-in-the-read-csv>

🡪 hm\_distribution2=read.csv("C:/Users/Legros Ferdinand/Documents/Pack Pricing/Data/Exploitable/FR\_Laundry\_HM\_Distribution.csv",header=TRUE, sep=";", na.strings="-", colClasses=c(rep("character",6),"numeric","character",rep("numeric",39)))

Convert data frame to vector by rows

<http://stackoverflow.com/questions/2545228/converting-a-dataframe-to-a-vector-by-rows>

Subset a data set rows : subset(frame,y>2)

<http://www.ats.ucla.edu/stat/r/faq/subset_R.htm>

Sommer les lignes d'un data frame

df$e <- rowSums(subset(df, select=a:d))

<http://stackoverflow.com/questions/9651202/sum-variables-within-r-dataframe>

Compter le nombre de valeur au dessus d'un nombre dans un vecteur

<http://stackoverflow.com/questions/3818147/count-number-of-vector-values-in-range-with-r>

Somme cumulée pour une colonne

apply(A, 2, cumsum)

Insert column in data frame

target <- which(names(d) == 'b')[1]

cbind(d[,1:target,drop=F], data.frame(d=12:15), d[,(target+1):length(d),drop=F])

<http://stackoverflow.com/questions/13502601/adding-a-column-between-two-columns-in-a-data-frame>

Sort a data frame according to column

**sort1.hsb2 <- hsb2[order(read) , ]**

<http://www.ats.ucla.edu/stat/r/faq/sort.htm>

Renommer la 2e colonne d'un dataset

colnames(trSamp)[2] <- "newname2"

Convertir une data frame en fichier csv (pour l'ouvrir avec Excel),

> write.table(test, "C:/Users/Legros Ferdinand/Documents/Pack Pricing/Data/Exploitable/R Traitee/test3.csv", col.names=NA, sep=" ; " )

Supprimer les lignes égales ou égales selon un critère contenu dans une colonne

Subset

Puis <http://stats.stackexchange.com/questions/6759/removing-duplicated-rows-from-r-data-frame>

Sommer certaines colonnes d'un data frame

colSums(dataset[,c(1,2)])

<http://r.789695.n4.nabble.com/Summing-Select-Columns-of-a-Data-Frame-td789884.html>

Renommer les colonnes d'un data frame

<http://www.cookbook-r.com/Manipulating_data/Renaming_columns_in_a_data_frame/>

Conseil de Julien: Programmer ses propres fonctions R pour ne pas risquer de commettre d'erreurs dans les copier-coller

Remplacer du text dans un string

group <- c("12357e", "12575e", "197e18", "e18947")

group

[1] "12357e" "12575e" "197e18" "e18947"

gsub("e", "", group)

[1] "12357" "12575" "19718" "18947"

<http://stackoverflow.com/questions/11936339/in-r-how-do-i-replace-text-within-a-string>

Itérer sur un nom de fichier

for(i in 0:3) {

infile <- paste(i,".txt",sep="")

outfile <- paste(i,"-edit.txt",sep="")

data <- read.table(infile,header=TRUE,sep=",",row.names=NULL)

colnames(data)[1] = "time"

write.table(data,quote=FALSE,sep=", ",outfile)

}

<http://stackoverflow.com/questions/4854130/how-to-iterate-over-file-names-in-a-r-script>

Splitter un string

test=data.frame(do.call('rbind', strsplit(sku,';',fixed=TRUE)))

<http://stackoverflow.com/questions/7069076/split-column-at-delimiter-in-data-frame>

Appliquer une function à une liste de data frames

Lapply

df.list <- list(df1,df2,...)

res <- lapply(df.list, function(x) rowMeans(subset(x, select = c(start, stop)), na.rm = TRUE))

# to keep the original data.frame also

res <- lapply(df.list, function(x) cbind(x,"rowmean"=rowMeans(subset(x, select = c(start, stop)), na.rm = TRUE)))

<http://stackoverflow.com/questions/22002838/same-function-over-multiple-data-frames-in-r>

Variable globale dans une function

a<<-2

Multiples Plot sauvegardés dans fichiers externes

df1 <- data.frame(var1=rnorm(10), var2=rnorm(10))

df2 <- data.frame(var1=rnorm(10), var2=rnorm(10))

comment(df1) <- "df1"

comment(df2) <- "df2"

for ( dataFrame in list(df1,df2) ) {

dFnm <- comment(dataFrame)

pdf(file=paste( dFnm, "\_var1\_vs\_var2.pdf", sep="" ))

plot( dataFrame[["var1"]], dataFrame[["var2"]] )

dev.off();

}

<http://stackoverflow.com/questions/10656758/how-to-convert-the-name-of-a-dataframe-to-a-string-in-r>

Passer les noms de data frame dans une boucle 🡪 passer le nom en comment

Comment(df1) <- "df1"

<http://stackoverflow.com/questions/10656758/how-to-convert-the-name-of-a-dataframe-to-a-string-in-r>

Créer un nom de variable avec "paste"

Assign(paste(…),5)

<http://stackoverflow.com/questions/5510966/create-a-variable-name-with-paste-in-r>

Assign global variable in function

assign("Global.res", x^2, envir = .GlobalEnv)

Plusieurs tracés sur un même plot

plot(x,y1,type="l",col="red")

lines(x,y2,col="green")

<http://stackoverflow.com/questions/2564258/plot-two-graphs-in-same-plot-in-r>

Ajouter une legend

legend(2000,9.5, # places a legend at the appropriate placec("Health","Defense"), # puts text in the legend   
  
lty=c(1,1), # gives the legend appropriate symbols (lines)  
  
lwd=c(2.5,2.5),col=c("blue","red")) # gives the legend lines the correct color and width

<http://learningrbasic.blogspot.fr/2009/07/adding-legend-to-plot.html>

SUMIFs

require(dplyr)

df <- group\_by(df, Date, SKU)

df.summary <- summarise(df,

sales = sum(Sales.in.USD),

units = sum(Sum.Units),

curr = sum(Sales.in.Loc.currency)

)

<http://stackoverflow.com/questions/24509165/sumif-in-r-multiple-criteria>

Passer en long format avec reshape

resLong <- **reshape**(dfWide, varying=**c**("DV\_t1", "DV\_t2", "DV\_t3"),

direction="long", idvar=**c**("id", "IVbtw"),

v.names="DV", timevar="IVwth")

resLong[**order**(resLong$id), ]

🡪 hm\_distributionLong <- reshape(hm\_distribution\_s, varying=c("X4.S.29.01.12", "X4.S.26.02.12", "X4.S.25.03.12", "X4.S.22.04.12", "X4.S.20.05.12"), direction="long", idvar=c("Retail", "SKU","Pack.2nd"), v.names="WD", timevar="Period")

**Deux variables sur le meme plot avec ggplot**

ggplot(test\_data, aes(date)) +

geom\_line(aes(y = var0, colour = "var0")) +

geom\_line(aes(y = var1, colour = "var1"))

<http://stackoverflow.com/questions/3777174/plotting-two-variables-as-lines-using-ggplot2-on-the-same-graph>

ou

ggplot(global\_ls, aes(Period, P.PNP.Volume.Mean, colour = Retail)) + geom\_line()

<http://stackoverflow.com/questions/11242420/plotting-two-variables-using-ggplot2-same-x-axis>

🡪 Couleurs different automatiques selon les variables

**Boîtes à moustaches – Boxplot**

ggplot(hm\_ggls.sku,aes(x=Retail,y=PNP.Price.Mean.Period)) + geom\_boxplot()

<http://www.r-bloggers.com/summarising-data-using-box-and-whisker-plots/>

**Histogramme à poids différend du cardinal**

ggplot(data=hm\_ggl.sku) + geom\_histogram(aes(x=NB.LAVAGES.UL,weight=PNP.Volume))

<http://stackoverflow.com/questions/25427734/how-to-plot-an-histogram-with-y-as-a-sum-of-the-x-values-for-every-bin-in-ggplot>

**Histograms and density curves**

<https://stat.ethz.ch/pipermail/r-help/2011-June/280588.html>

<http://stackoverflow.com/questions/3541713/how-to-plot-two-histograms-together-in-r>

**GGplot title font**

bp + ggtitle("Plant growth with\ndifferent treatments") +

theme(plot.title = element\_text(lineheight=.8, face="bold"))

<http://www.cookbook-r.com/Graphs/Titles_(ggplot2)/>

**Avoir la range = (min,max) d'un vecteur de valeurs**

range(df$value)

range(df$value)[1] 🡪 min

range(df$value)[2] 🡪 max

**Passer le nom d'une colonne en argument d'une fonction**

<http://stackoverflow.com/questions/2641653/pass-a-data-frame-column-name-to-a-function>

**Histogram where wieght = mean**

<http://stackoverflow.com/questions/13408503/histogram-of-mean-per-bin>

**Deux data frames différents pour un même plot**

<http://stackoverflow.com/questions/9109156/ggplot-combining-two-plots-from-different-data-frames>

**Taille des xlabel et ylabel pour ggplot2**

<http://stackoverflow.com/questions/14942681/change-size-of-axes-title-and-labels-in-ggplot2>

**Tuto pour débutants en R**

<http://www.computerworld.com/article/2486425/business-intelligence/business-intelligence-4-data-wrangling-tasks-in-r-for-advanced-beginners.html>

**On n'utilise pas summary() dans une custom function**

<http://www.dummies.com/how-to/content/how-to-summarize-a-dataset-in-r.html>

**Enregistrer un plot via code**

fit <- lm(some ~ model)

png(

"test.png",

width = 3.25,

height = 3.25,

units = "in",

res = 300,

pointsize = 4

)

plot(fit)

dev.off()

<http://stackoverflow.com/questions/7144118/how-to-save-a-plot-as-image-on-the-disk>

<http://stackoverflow.com/questions/8399100/r-plot-size-and-resolution>

**Enregistrer un ggplot via code**

🡪 Fonction ggsave

**Grille de plot ggplot2**

🡪 Fonction grid.arrange, package gridExtra

Liste avec tous les data frames de l'environnement

dfs <- Filter(function(x) is(x, "data.frame"), mget(ls()))

<http://stackoverflow.com/questions/25509879/how-can-i-make-a-list-of-all-dataframes-that-are-in-my-global-environment>

**Introduire les items d'une liste dans le global environment**

> obj <- list(a=1:5, b=2:10, c=-5:5)

> ls()

[1] "obj"

> list2env(obj, .GlobalEnv)

<environment: R\_GlobalEnv>

> ls()

[1] "a" "b" "c" "obj"

<http://stackoverflow.com/questions/13795526/return-elements-of-r-list-as-independent-objects-in-global-environment>

**Auto indentation**

Ctrl + Shift + A

**Utiliser une fonction apply avec des plusieurs paramètres**

hello=function(x,parameter) {…}

apply(df,function,parameter=1)

<http://stackoverflow.com/questions/14427253/passing-several-arguments-to-fun-of-lapply-and-others-apply>

**Supprimer par nom un élément d'une liste**

<http://stackoverflow.com/questions/16778161/remove-elements-of-a-list-explicitly>

**Supprimer des colonnes d'un data frame par nom**

R> subset(df, select=-c(z,u))

x y

1 1 2

2 2 3

3 3 4

4 4 5

5 5 6

<http://stackoverflow.com/questions/5234117/how-to-drop-columns-by-name-in-a-data-frame>

**Tester si une liste contient un certain élément**

is.null(liste[["a"]])

<http://stackoverflow.com/questions/7719741/how-to-test-if-list-element-exists>

**Matrice de correlation**

corr(df, use = "…")

**Plots croisés**

pairs ()

Aussi possible avec ggplot

**Regression polynomiale**

<http://www.r-bloggers.com/polynomial-regression-techniques/>

**Regression linéaire par morceaux**

<http://www.r-bloggers.com/r-for-ecologists-putting-together-a-piecewise-regression/>

<http://www.r-bloggers.com/estimating-continuous-piecewise-linear-regression/>

**Tutos rpart**

<http://www.statmethods.net/advstats/cart.html>

<http://www.r-bloggers.com/a-brief-tour-of-the-trees-and-forests/>

<http://www.edureka.co/blog/implementation-of-decision-tree/> ++

**Definitions rpart**

cp de printcp: amount by which splitting that node would decrease the relative error

Relative error: 1−R² Root mean squared error, similar to linear regression

xerror: error on the observations from cross validation data

<http://stackoverflow.com/questions/9666212/how-to-compute-error-rate-from-a-decision-tree>

<http://www.link.ch/wp-content/uploads/2013/10/CrossValidationFinal.pdf>

🡪 Question: definition de "Root node error"?

xstd

**Mélanger les lignes / colonnes d'un data frame**

🡪 Sample

df2 <- df1[sample(nrow(df1)),]

<http://stackoverflow.com/questions/6422273/how-to-randomize-or-permute-a-dataframe-rowwise-and-columnwise>

**Cross Validation pour random forest**

🡪 Pas forcément nécessaire !

**Test sur des données de random forest**

library(randomForest)

test <- iris[ c(1:10, 51:60, 101:110), -5] # removed the Species column here.

train <- iris[ c(11:50, 61:100, 111:150), ]

r <- randomForest(Species ~., data=train, importance=TRUE, do.trace=100)

predict(r, test)

<http://stackoverflow.com/questions/17479632/random-forests-does-it-make-any-difference-if-the-test-set-is-also-labeled>

**Visualisation d'une regression multiple**

<http://stats.stackexchange.com/questions/89747/how-to-describe-or-visualize-a-multiple-linear-regression-model>

+++

**Executer un string comme commande**

a = 1

b = 2

c = parse(text = "log(a + b)")

eval(c)

**Raccourcis de code R studio**

<https://support.rstudio.com/hc/en-us/articles/200484568-Code-Folding-and-Sections>

**Premier TRUE d'un vecteur booléen**

min(which(lv == TRUE))

**Visualisation d'une matrice de corrélation**

<http://www.sthda.com/english/wiki/correlation-matrix-a-quick-start-guide-to-analyze-format-and-visualize-a-correlation-matrix-using-r-software>

++