ggwordcloud.R

ikapratiwi

Fri Nov 27 14:32:09 2015

@import tm @importFrom grid unit

wcloudGrob <- function (label,x = unit(0.5, "npc"), y = unit(0.5, "npc"),  
 just = "centre", hjust = NULL, vjust = NULL, rot = 0, check.overlap = TRUE,  
 default.units = "mm", name = NULL, gp = gpar(), vp = NULL, f=7.62) {  
 if (!requireNamespace("tm", quietly = TRUE)) {  
 stop("Pkg needed for this function to work. Please install it.",  
 call. = FALSE)  
 }  
 if (is.character(label) || is.factor(label)) {  
 corpus <- Corpus(VectorSource(label))  
 corpus <- tm\_map(corpus, removePunctuation)  
 #corpus <- tm\_map(corpus, removeNumbers)  
 corpus <- tm\_map(corpus, function(x) removeWords(x,stopwords()))  
 } else corpus <- label  
 tdm <- TermDocumentMatrix(corpus)  
 mat<- as.matrix(tdm)  
 v <- sort(rowSums(mat),decreasing=TRUE)  
 word <- names(v)  
 freq <- v  
  
 a <- 4  
 b <- 0.5  
 scales <- freq/max(freq)  
 sizes <- ((a-b)\*scales+b)  
  
 loc<-matrix(nrow=length(word),ncol=2)  
 ######created by Ian Fellows  
 boxes <- list()  
 thetaStep <- 0.1  
 rStep <- 0.05  
 last <- 1  
 tails <- "g|j|p|q|y"  
 overlap <- function(x1, y1, sw1, sh1) {  
 s <- 0  
 if (length(boxes) == 0)  
 return(FALSE)  
 for (i in c(last, 1:length(boxes))) {  
 bnds <- boxes[[i]]  
 x2 <- bnds[1]  
 y2 <- bnds[2]  
 sw2 <- bnds[3]  
 sh2 <- bnds[4]  
 if (x1 < x2)  
 overlap <- x1 + sw1 > x2 - s  
 else overlap <- x2 + sw2 > x1 - s  
 if (y1 < y2)  
 overlap <- overlap && (y1 + sh1 > y2 - s)  
 else overlap <- overlap && (y2 + sh2 > y1 - s)  
 if (overlap) {  
 last <<- i  
 return(TRUE)  
 }  
 }  
 FALSE  
 }  
  
 for(i in 1:length(sizes)){  
 r <- 0  
 theta <- runif(1, 0, 2 \* pi)  
 x1 <- 0.5  
 y1 <- 0.5  
 w <- strwidth(word[i],cex = sizes[i],"inches")/5.08 #I modified this to match the unit measurement for ggplot  
 h <- strheight(word[i],cex = sizes[i],"inches")/5.08 #I modified this to match the unit measurement for ggplot  
 if (grepl(tails, word[i]))  
 h <- h + h \* 0.2  
 isOverlaped <- TRUE  
 while (isOverlaped) {  
 if (!overlap(x1 - 0.5 \* w, y1 - 0.5 \* h, w, h) && x1 - 0.5 \* w > 0 && y1 - 0.5 \* h > 0 && x1 + 0.5 \* w < 1 && y1 + 0.5 \* h < 1) {  
 boxes[[length(boxes) + 1]] <- c(x1 - 0.5 \* w,  
 y1 - 0.5 \* h, w, h)  
 loc[i,] <- c(x1,y1) #I modified this to store the coordinates that are not overlapped  
 isOverlaped <- FALSE  
 } else  
 {  
 if (r > sqrt(0.5)) {  
 isOverlaped <- FALSE  
 }  
 theta <- theta + thetaStep  
 r <- r + (rStep \* thetaStep/(2 \* pi))  
 x1 <- 0.5 + r \* cos(theta)  
 y1 <- 0.5 + r \* sin(theta)  
 }  
 }  
 }  
 ######  
  
 x <- unit(loc[,1], default.units)  
 y <- unit(loc[,2], default.units)  
  
 grob(label = word, x = x, y = y, just = just, hjust = hjust,  
 vjust = vjust, rot = rot, check.overlap = check.overlap,  
 name = name, gp = gp, vp = vp, cl = "text")  
 tg <- textGrob(label = word, x = x, y = y, just = just, hjust = hjust,  
 vjust = vjust, rot = rot, check.overlap = check.overlap, gp=gpar(cex=sizes))  
  
 gTree(children=gList(tg), vp=vp, gp=gp, name=name)  
}

@import proto

GeomWordCloud <- proto::proto(ggplot2:::GeomText, {  
 objname <- "word\_cloud"  
  
 draw <- function(., data, scales, coordinates, ..., parse = FALSE, na.rm = FALSE) {  
 data <- remove\_missing(data, na.rm,  
 c("label"), name = "geom\_word\_cloud")  
  
 lab <- data$label  
  
 with(coord\_transform(coordinates, data, scales),  
 wcloudGrob(lab, x, y, default.units="native",  
 hjust=hjust, vjust=vjust, rot=angle,  
 gp = gpar(col = alpha(colour, alpha),  
 fontfamily = family, fontface = fontface, lineheight = lineheight))  
 )  
 }  
})

Create a new geom called wordcloud

@param data data frame contains the words to visualize @details Set aes(x=0, y=0) as a default if the data frame only contains the words @examples words2 <- "South Korea high-energy capital moves at a pace that rivals the world busiest, with its pulsing creative energy driving innovation in music, technology and fashion.Pali pali ??? meaning quick, quick ??? is not just a favourite expression in Seoul: it is a way of life. South Korea capital moves at a pace that rivals the world's busiest cities, fostering a culture of hard work, service and just getting stuff done, residents say. It is the most high-energy, intense city I have ever lived in. I have lived in New York City and Tokyo, but Seoul beats them hands down, said Ruchika Sahai, who moved here from Sydney, Australia, nearly two years ago. Office hours stretch well past 10 pm, but the drive of the nearly 10 million residents rarely wavers. Even beyond the modern city streets, the underground markets pulse with a chaotic and old world vibe" dats <- data.frame(words2) qplot(0,0,data=dats,label=dats$words2) +geom\_word\_cloud(colour="red")

p1 <- ggplot(data=dats,aes(x=0,y=0,label=dats$words2)) p1+geom\_word\_cloud(colour="red")

p <- ggplot(data = mtcars, aes(x=wt, y=mpg, label=rownames(mtcars))) p+geom\_word\_cloud()

# Add aesthetic mappings

p+geom\_word\_cloud(aes(colour=factor(cyl))) p+geom\_word\_cloud(aes(colour=factor(cyl)))+scale\_colour\_discrete(l=40)

geom\_word\_cloud <- function (mapping = NULL, data = NULL, stat = "identity", position = "identity",  
 parse = FALSE, ...) {  
 GeomWordCloud$new(mapping = mapping, data = data, stat = stat,position = position,  
 parse = parse, ...)  
}