correlation plot

Stephanie Yang jy2777

12/6/2017

```
library("corrplot")
```

```
## corrplot 0.84 loaded
```

```
setwd("~/Desktop/Homework/Statistical Methods/Project/datasets")
#transpose t()
starbucks.metrics <- t(read.csv("starbucks_metrics2.csv", header=FALSE))

#labels columns
colnames(starbucks.metrics)<- c("Date","Likes (Total) FB","Comments (Total) FB","Shares (Total) FB","Rea

#removes duplicate row
starbucks.metrics1 <- starbucks.metrics[-1,]

##Cleaning the metrics sheet
#1) removes space in column titles
colnames(starbucks.metrics1) <- gsub(" ","",colnames(starbucks.metrics1))

#2) removes % symbol of column 10
starbucks.metrics1[,c(11,15,19,20,21,26,27,28,31,33,38,39,43,44,52,59,60,61,69,71,72,73)] <- as.numeric
starbucks.metrics1 <- as.data.frame(starbucks.metrics1)
class(starbucks.metrics1)
```

```
## [1] "data.frame"
```

```
#3) removes comma separator for thousands, except for date column which is type character not numeric
#gsub to replace "," with "", and then convert the string to numeric using as.numeric
starbucks.metrics1[,2:73] <- lapply(starbucks.metrics1[,2:73], function(x) as.numeric(gsub(",","", as.cl
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
##Transforms Monthly to Quarterly Data:
library("lubridate")
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
#creates a year and quarter column per row
starbucks.metrics1$Date <- ymd(starbucks.metrics1$Date)
```

```
## Warning in as.POSIXlt.POSIXct(x, tz): unknown timezone 'zone/tz/2017c.1.0/
## zoneinfo/America/New_York'
```

```
starbucks.metrics1$year = year(starbucks.metrics1$Date)
starbucks.metrics1$quarter = quarter(starbucks.metrics1$Date)

#aggregates quarters of same year and takes their sum (sales are aldo sums) : HOW TO DO IT WITH AGGREGA
```

```
library("reshape2")
starbucks.metrics2 <- melt(starbucks.metrics1[,2:75], id=c("quarter", "year"))
starbucks.metrics2 <- dcast(starbucks.metrics2, year + quarter ~ variable, fun.aggregate = sum)
write.csv(starbucks.metrics2,file="colgate_vizmetrics.csv")
starbucks.metrics2 <- starbucks.metrics2[1:9,]

starbucks.metrics3 <- starbucks.metrics2[,-c(1,2)]
starbucks.metrics3 <- starbucks.metrics3[complete.cases(starbucks.metrics3),]

starbucks.metrics4 <- starbucks.metrics3[,-c(15,45)]



corr.starbuck <- cor(starbucks.metrics4)

starbucks.corr <- corrplot(corr.starbuck, type="full", order="hclust", tl.col="black", tl.srt=0.01,tl.c
```