PCA

Stephanie Yang jy2777

12/3/2017

PCA We did PCA seperately for all 16 companies and here is an example of Chipotle.

```
setwd("~/Desktop/Homework/Statistical Methods/Project/datasets")
#transpose t()
chipotle.metrics <- t(read.csv("chipotle_metrics22.csv", header=FALSE))

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

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

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

#2) removes % symbol of column 10
chipotle.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(
chipotle.metrics1 <- as.data.frame(chipotle.metrics1)
class(chipotle.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
chipotle.metrics1[,2:73] <- lapply(chipotle.metrics1[,2:73], function(x) as.numeric(gsub(",","", as.cha
```

```
## 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
chipotle.metrics1$Date <- ymd(chipotle.metrics1$Date)
```

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

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

#aggregates quarters of same year and takes their sum (sales are aldo sums) : HOW TO DO IT WITH AGGREGA
library("reshape2")
chipotle.metrics2 <- melt(chipotle.metrics1[,2:75], id=c("quarter", "year"))
```

```
chipotle.metrics2 <- dcast(chipotle.metrics2, year + quarter ~ variable, fun.aggregate = sum)

#aggregate(chipotle.metrics1$year ~.,chipotle.metrics1, FUN = mean, na.action = na.omit)
write.csv(chipotle.metrics2,file="chipotle_vizmetrics.csv")

chipotle.metrics3 <- chipotle.metrics2[,-c(1:2)]
chipotle.metrics3 <- chipotle.metrics3[complete.cases(chipotle.metrics3),]
output.chi <- prcomp(chipotle.metrics3)
output.chi$sdev[1]^2/sum(output.chi$sdev^2)
```

```
## [1] 0.8363353
```

```
output.chi$sdev[2]^2/sum(output.chi$sdev^2)
```

```
## [1] 0.1258793
```

```
output.chi$sdev[3]^2/sum(output.chi$sdev^2)
```

```
## [1] 0.03680482
```