```
###### 1.import data #####
apps <- read.csv("/Users/ninenins/Desktop/R/Final project Yang/AppleStore
7200.csv",
                  stringsAsFactors = FALSE,na.strings = "")
summary(apps)
###### 2.basic description of 4 parts #####

## type
type <- apps$prime_genre
x <- table(type)
pie(x,main="Apps Type")

## size(transfer from bytes to MB)
size <- apps$size_bytes
sizeMB <- size * 10**(-6)
apps <- cbind(apps,sizeMB)
sizerank <- apps[order(sizeMB,decreasing = T),]
head(sizerank)
q <- quantile(sizeMB)

size1 <- length(apps$sizeMB[apps$sizeMB<= 50])
size2 <- length(apps$sizeMB[apps$sizeMB > 50 & apps$sizeMB<= 100])
size3 <- length(apps$sizeMB[apps$sizeMB > 100 & apps$sizeMB<= 200])
size4 <- length(apps$sizeMB[apps$sizeMB> 200])
totalsize <- c(size1,size2,size3,size4)
barplot(totalsize,main = "Size",names= c("<50", "50~100", "100~200",
">200"))

## price
price <- apps$price
barplot(table(price),col = "blue",main = "Price($)")
summary(price)

## rating
rate <- apps$user_rating
summary(rate)
barplot(table(rate),col = "red",main = "Rating value")

##### 3.PCA #####
#comparing 1~100 social networking apps#
newapps <- split(apps,apps$prime_genre)
social <- newapps$`Social
Networking`[,c("track_name","sizeMB","price","user_rating","user_rating_ver",

                          "lang.num","sup_devices.num","ipadSc_urls.num")]

social.100 <- social[1:100,2:6]
pairs(social.100,main="social.100")
rownames(social.100) <- social[1:100,1]
head(social.100)

pc.social.100 <- prcomp(social.100 , scale.=TRUE)
pc.social.100$rotation
```

```
summary(pc.social.100)
plot(pc.social.100, type="l")
biplot(pc.social.100)


##### 4.EDA #####
cor.apps <-  cor(apps[,c("sizeMB","price","user_rating","rating_count_tot",
                         "lang.num","sup_devices.num","ipadSc_urls.num")])
round(cor.apps,3)
cor.apps[lower.tri(cor.apps,diag=TRUE)] = 0
cor.apps.sorted = sort(abs(cor.apps),decreasing=T)
cor.apps.sorted[cor.apps.sorted>0]
a <- which(abs(cor.apps)==cor.apps.sorted[1])
a
var.big.cor <- arrayInd(a, dim(cor.apps))
var.big.cor1 <- c(a%%nrow(cor.apps), (a%/%nrow(cor.apps))+1)
colnames(cor.apps)[var.big.cor]



##### regression ####
apps$rating_count_tot <- apps$rating_count_tot*0.001
reg <- lm(apps$user_rating~apps$sizeMB+apps$price+apps$sup_devices.num
          +apps$ipadSc_urls.num+apps$lang.num+apps$rating_count_tot)
summary(reg)
## call coefficient
reg.coef <- coef(reg)
reg.coef
round(reg$coefficients,5)
plot(reg)
```