101a final project

Happiness <- read.delim("~/Downloads/Happiness.txt")  
  
Happiness$RowNumber <- seq.int(nrow(Happiness))  
  
attach(Happiness)  
sqrtTransform <- lm(Happy~sqrt(Household)+  
 sqrt(Age)+  
 sqrt(Children)+  
 sqrt(OwnHome)+  
 sqrt(Instagram)+  
 sqrt(Marital) +  
 sqrt(Sex)+  
 sqrt(Education)+  
 sqrt(JobSat)+  
 sqrt(Income)+  
 sqrt(WorkHrs))

## Warning in sqrt(WorkHrs): NaNs produced

summary(sqrtTransform)

##   
## Call:  
## lm(formula = Happy ~ sqrt(Household) + sqrt(Age) + sqrt(Children) +   
## sqrt(OwnHome) + sqrt(Instagram) + sqrt(Marital) + sqrt(Sex) +   
## sqrt(Education) + sqrt(JobSat) + sqrt(Income) + sqrt(WorkHrs))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.5216 -0.5595 0.0244 0.3621 7.3182   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.3249630 0.4882234 0.666 0.50599   
## sqrt(Household) -0.2214753 0.1295527 -1.710 0.08801 .   
## sqrt(Age) 0.0050998 0.0360172 0.142 0.88746   
## sqrt(Children) 0.0948532 0.0500377 1.896 0.05862 .   
## sqrt(OwnHome) NA NA NA NA   
## sqrt(Instagram) -0.0126475 0.0625477 -0.202 0.83984   
## sqrt(Marital) 0.1944156 0.0764610 2.543 0.01132 \*   
## sqrt(Sex) 0.2581200 0.1726004 1.495 0.13546   
## sqrt(Education) 0.1312890 0.0593076 2.214 0.02733 \*   
## sqrt(JobSat) 0.4362079 0.0790955 5.515 5.74e-08 \*\*\*  
## sqrt(Income) -0.0014482 0.0004599 -3.149 0.00174 \*\*   
## sqrt(WorkHrs) 0.0166643 0.0075174 2.217 0.02711 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7491 on 474 degrees of freedom  
## (1882 observations deleted due to missingness)  
## Multiple R-squared: 0.1427, Adjusted R-squared: 0.1246   
## F-statistic: 7.889 on 10 and 474 DF, p-value: 8.988e-12

Happiness$Household[Happiness$Household==9] <- NA  
#Age  
Happiness$Age[Happiness$Age==99] <- NA  
#OwnHome  
Happiness$OwnHome[Happiness$OwnHome==8 | Happiness$OwnHome==9] <- NA  
Happiness$OwnHome[Happiness$OwnHome==0] <- 4  
#Marital  
Happiness$Marital[Happiness$Marital==9] <- NA  
#JobSat  
Happiness$JobSat[which(Happiness$JobSat==9)]<-NA  
Happiness$JobSat[which(Happiness$JobSat==0)]<-9  
#Happy  
Happiness$Happy[Happiness$Happy == 0|Happiness$Happy==8|Happiness$Happy==9] <- NA  
#Health  
Happiness$Health[which(Happiness$Health==9)]<-NA  
Happiness$Health[which(Happiness$Health==0)]<-9  
#Instagram  
Happiness$Instagram[Happiness$Instagram==0] <- 3  
#Education  
Happiness$Education[Happiness$Education == 97 | Happiness$Education == 98 | Happiness$Education == 99] <- NA  
#Income  
Happiness$Income[Happiness$Income == 0 | Happiness$Income == 999998 | Happiness$Income == 999999] <- NA  
#WorkHrs  
Happiness$WorkHrs[Happiness$WorkHrs==-1]<-NA  
finaldata <- na.omit(Happiness)  
  
#basic  
m1 <- lm(Happiness$Happy~Happiness$Household+Happiness$Health+Happiness$OwnHome+Happiness$Instagram+Happiness$Marital+Happiness$Sex+Happiness$Age+Happiness$Children+Happiness$Education+Happiness$JobSat+Happiness$Income+Happiness$WorkHrs, na.action=na.omit)  
summary(m1)

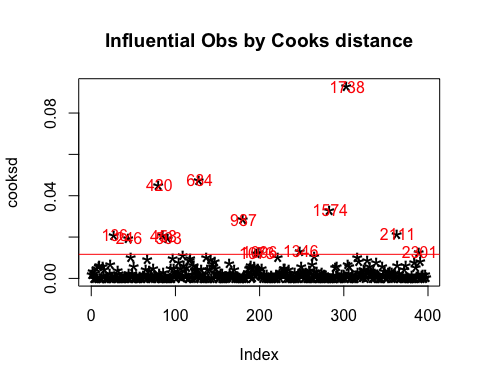
##   
## Call:  
## lm(formula = Happiness$Happy ~ Happiness$Household + Happiness$Health +   
## Happiness$OwnHome + Happiness$Instagram + Happiness$Marital +   
## Happiness$Sex + Happiness$Age + Happiness$Children + Happiness$Education +   
## Happiness$JobSat + Happiness$Income + Happiness$WorkHrs,   
## na.action = na.omit)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.29952 -0.56558 0.08724 0.36169 1.50048   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.506e-01 2.967e-01 3.204 0.001470 \*\*   
## Happiness$Household -7.704e-02 3.999e-02 -1.926 0.054822 .   
## Happiness$Health 1.475e-01 4.409e-02 3.345 0.000904 \*\*\*  
## Happiness$OwnHome NA NA NA NA   
## Happiness$Instagram 4.032e-02 3.867e-02 1.043 0.297767   
## Happiness$Marital 6.603e-02 2.157e-02 3.061 0.002359 \*\*   
## Happiness$Sex 1.164e-01 6.485e-02 1.795 0.073365 .   
## Happiness$Age -3.758e-04 2.573e-03 -0.146 0.883967   
## Happiness$Children 2.420e-02 2.221e-02 1.089 0.276661   
## Happiness$Education 1.040e-02 1.102e-02 0.944 0.345774   
## Happiness$JobSat 5.521e-02 2.243e-02 2.462 0.014261 \*   
## Happiness$Income -9.896e-07 1.318e-06 -0.751 0.453024   
## Happiness$WorkHrs -2.456e-05 2.197e-04 -0.112 0.911051   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6118 on 386 degrees of freedom  
## (1969 observations deleted due to missingness)  
## Multiple R-squared: 0.1208, Adjusted R-squared: 0.09578   
## F-statistic: 4.823 on 11 and 386 DF, p-value: 5.567e-07

#Cooks distance  
cooksd <- cooks.distance(m1)  
plot(cooksd, pch="\*", cex=2, main="Influential Obs by Cooks distance")  
abline(h = 4\*mean(cooksd, na.rm=T), col="red")  
text(x=1:length(cooksd)+1, y=cooksd, labels=ifelse(cooksd>4\*mean(cooksd, na.rm=T),names(cooksd),""), col="red")  
  
#inverse response  
library(quantreg)

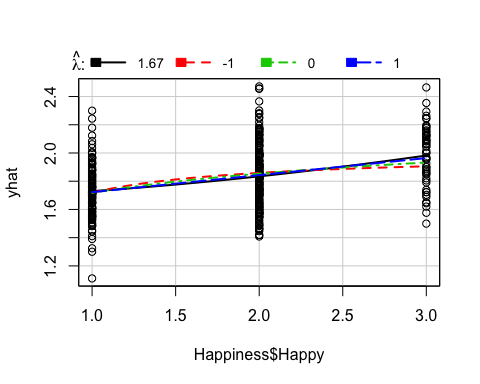
## Loading required package: SparseM

##   
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':  
##   
## backsolve



library(car)  
library(alr3)  
par(mfrow=c(1,1))  
inverseResponsePlot(m1,key=TRUE)

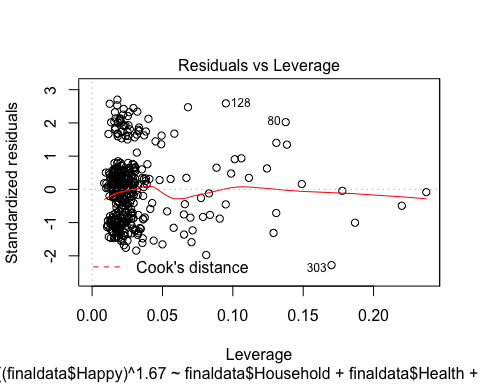
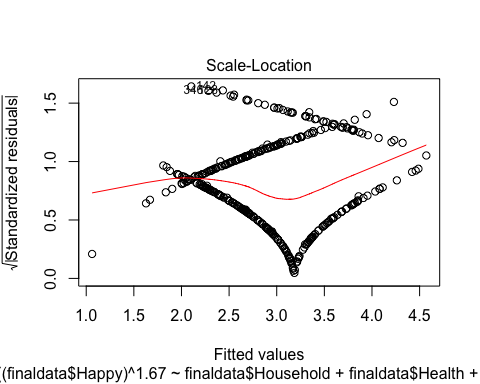
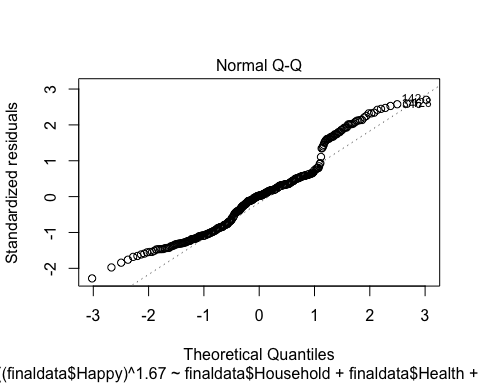
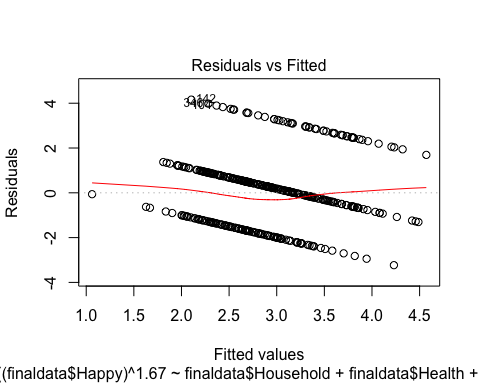


## lambda RSS  
## 1 1.670272 17.41699  
## 2 -1.000000 17.87376  
## 3 0.000000 17.64331  
## 4 1.000000 17.45815

m2 <- lm((finaldata$Happy)^1.67~finaldata$Household+finaldata$Health+finaldata$Instagram+finaldata$Marital+finaldata$Sex+finaldata$Age+finaldata$Children+finaldata$Education+finaldata$JobSat+finaldata$Income+finaldata$WorkHrs)  
summary(m2)

##   
## Call:  
## lm(formula = (finaldata$Happy)^1.67 ~ finaldata$Household + finaldata$Health +   
## finaldata$Instagram + finaldata$Marital + finaldata$Sex +   
## finaldata$Age + finaldata$Children + finaldata$Education +   
## finaldata$JobSat + finaldata$Income + finaldata$WorkHrs)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.2303 -1.2875 0.0390 0.7782 4.1598   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.534e-01 7.540e-01 0.999 0.3183   
## finaldata$Household -2.100e-01 1.016e-01 -2.066 0.0395 \*   
## finaldata$Health 3.656e-01 1.120e-01 3.264 0.0012 \*\*  
## finaldata$Instagram 1.078e-01 9.826e-02 1.097 0.2735   
## finaldata$Marital 1.742e-01 5.482e-02 3.178 0.0016 \*\*  
## finaldata$Sex 3.248e-01 1.648e-01 1.971 0.0494 \*   
## finaldata$Age -9.211e-04 6.539e-03 -0.141 0.8880   
## finaldata$Children 7.482e-02 5.644e-02 1.326 0.1858   
## finaldata$Education 2.001e-02 2.801e-02 0.714 0.4755   
## finaldata$JobSat 1.300e-01 5.699e-02 2.281 0.0231 \*   
## finaldata$Income -2.386e-06 3.348e-06 -0.713 0.4764   
## finaldata$WorkHrs 1.296e-04 5.582e-04 0.232 0.8166   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.555 on 386 degrees of freedom  
## Multiple R-squared: 0.1237, Adjusted R-squared: 0.09874   
## F-statistic: 4.954 on 11 and 386 DF, p-value: 3.271e-07

plot(m2)



#doesn't change diagnostic plots much  
  
##log transformation with no children  
Happiness$Education[Happiness$Education == 0] <- NA  
finaldata <- na.omit(Happiness)  
m4 <- lm(log(Happy)~log(Household)+log(Health)+log(Instagram)+log(Marital)+  
 log(Sex)+log(Age)+log(JobSat)+log(Education)+log(Income)+log(WorkHrs), data=finaldata)  
  
#Joy  
  
attach(finaldata)

## The following objects are masked from Happiness:  
##   
## Age, Children, Education, Happy, Health, Household, Income,  
## Instagram, JobSat, Marital, OwnHome, RowNumber, Sex, WorkHrs

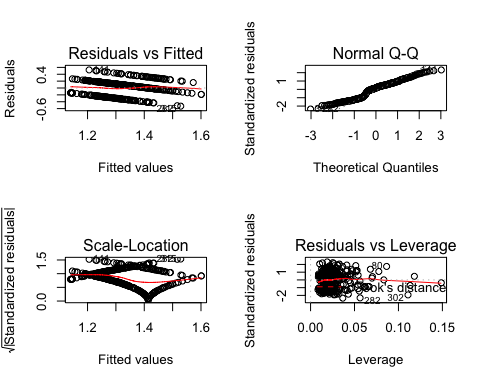
#Boxcox  
summary(tranxy <- powerTransform(cbind(Happy,Household,Age,Instagram,Marital,Sex,JobSat,Income,WorkHrs,Health)~1))

## bcPower Transformations to Multinormality   
## Est Power Rounded Pwr Wald Lwr bnd Wald Upr Bnd  
## Happy 0.6271 0.50 0.3541 0.9000  
## Household -0.1947 0.00 -0.4096 0.0203  
## Age 0.5542 0.50 0.2240 0.8845  
## Instagram 0.1340 0.00 -0.0464 0.3144  
## Marital -0.2480 -0.33 -0.4559 -0.0400  
## Sex 0.4587 0.00 -0.0342 0.9515  
## JobSat -0.0256 0.00 -0.1829 0.1316  
## Income 0.1940 0.19 0.1312 0.2569  
## WorkHrs -0.2203 -0.22 -0.2756 -0.1651  
## Health 0.5020 0.50 0.2639 0.7400  
##   
## Likelihood ratio tests about transformation parameters  
## LRT df pval  
## LR test, lambda = (0 0 0 0 0 0 0 0 0 0) 162.0843 10 0  
## LR test, lambda = (1 1 1 1 1 1 1 1 1 1) 2518.7490 10 0

aMarital<-Marital^(-1/3)  
aJobSat<-JobSat^(1/3)  
aIncome<-Income^(0.19)  
aWorkhrs<-WorkHrs^(-0.23)  
  
#Square root  
m2<-lm(sqrt(Happy)~log(Household)+sqrt(Age)+log(Instagram)+aMarital+log(Sex)+aJobSat+aIncome+aWorkhrs+sqrt(Health))  
  
summary(m2)

##   
## Call:  
## lm(formula = sqrt(Happy) ~ log(Household) + sqrt(Age) + log(Instagram) +   
## aMarital + log(Sex) + aJobSat + aIncome + aWorkhrs + sqrt(Health))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.52860 -0.22822 0.03974 0.15934 0.52650   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.062401 0.184284 5.765 1.67e-08 \*\*\*  
## log(Household) -0.086859 0.031731 -2.737 0.00648 \*\*   
## sqrt(Age) -0.007127 0.012807 -0.557 0.57818   
## log(Instagram) 0.026321 0.030744 0.856 0.39245   
## aMarital -0.169448 0.075632 -2.240 0.02563 \*   
## log(Sex) 0.054593 0.035187 1.552 0.12159   
## aJobSat 0.173040 0.052714 3.283 0.00112 \*\*   
## aIncome -0.007826 0.011340 -0.690 0.49054   
## aWorkhrs 0.202225 0.204077 0.991 0.32234   
## sqrt(Health) 0.138968 0.044619 3.115 0.00198 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2286 on 387 degrees of freedom  
## Multiple R-squared: 0.1407, Adjusted R-squared: 0.1208   
## F-statistic: 7.043 on 9 and 387 DF, p-value: 1.891e-09

par(mfrow=c(2,2))  
plot(m2)



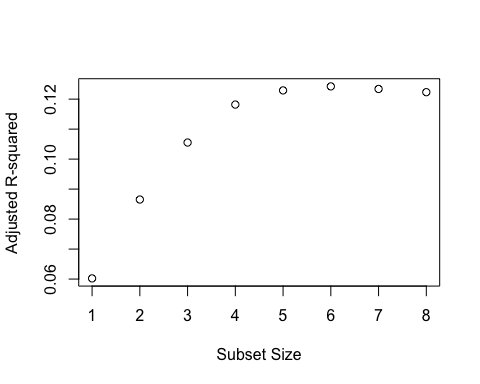
detach(finaldata)  
  
#All possible subsets  
  
finaldata <- na.omit(Happiness)  
bMarital<- finaldata$Marital^(-1/3)  
bJobSat<- finaldata$JobSat^(1/3)  
bIncome<- finaldata$Income^(0.19)  
bWorkhrs<-finaldata$WorkHrs^(-0.23)  
library(leaps)  
X <- cbind(log(finaldata$Household),sqrt(finaldata$Age), log(finaldata$Instagram), bMarital, log(finaldata$Sex), bJobSat, bIncome, bWorkhrs, sqrt(finaldata$Health))  
b <- regsubsets(as.matrix(X), sqrt(finaldata$Happy))  
rs <- summary(b)  
rs$adjr2

## [1] 0.06022350 0.08652574 0.10553126 0.11819139 0.12291154 0.12424638  
## [7] 0.12339516 0.12232015

rs

## Subset selection object  
## 9 Variables (and intercept)  
## Forced in Forced out  
## FALSE FALSE  
## FALSE FALSE  
## FALSE FALSE  
## bMarital FALSE FALSE  
## FALSE FALSE  
## bJobSat FALSE FALSE  
## bIncome FALSE FALSE  
## bWorkhrs FALSE FALSE  
## FALSE FALSE  
## 1 subsets of each size up to 8  
## Selection Algorithm: exhaustive  
## bMarital bJobSat bIncome bWorkhrs   
## 1 ( 1 ) " " " " " " "\*" " " " " " " " " " "  
## 2 ( 1 ) " " " " " " "\*" " " "\*" " " " " " "  
## 3 ( 1 ) "\*" " " " " " " " " "\*" " " " " "\*"  
## 4 ( 1 ) "\*" " " " " "\*" " " "\*" " " " " "\*"  
## 5 ( 1 ) "\*" " " " " "\*" "\*" "\*" " " " " "\*"  
## 6 ( 1 ) "\*" " " " " "\*" "\*" "\*" " " "\*" "\*"  
## 7 ( 1 ) "\*" " " " " "\*" "\*" "\*" "\*" "\*" "\*"  
## 8 ( 1 ) "\*" " " "\*" "\*" "\*" "\*" "\*" "\*" "\*"

par(mfrow=c(1,1))  
plot(1:8,rs$adjr2,xlab="Subset Size",ylab="Adjusted R-squared")



#According to the plot, the optimal subset of predictors of size 6 (log(Household), bMarital, log(Sex), bJobSat, bWorkhrs, sqrt(finaldata$Health)) maximizes adjusted R-square. Let's check other criteria for the optimal models.  
  
om1 <- lm(sqrt(finaldata$Happy)~bMarital)  
om2 <- lm(sqrt(finaldata$Happy)~bMarital+bJobSat)  
om3 <- lm(sqrt(finaldata$Happy)~log(finaldata$Household)+bJobSat+sqrt(finaldata$Health))  
om4 <- lm(sqrt(finaldata$Happy)~log(finaldata$Household)+bMarital+bJobSat+sqrt(finaldata$Health))  
om5 <- lm(sqrt(finaldata$Happy)~log(finaldata$Household)+bMarital+log(finaldata$Sex)+bJobSat+sqrt(finaldata$Health))  
om6 <- lm(sqrt(finaldata$Happy)~log(finaldata$Household)+bMarital+log(finaldata$Sex)+bJobSat+bWorkhrs+sqrt(finaldata$Health))  
om7 <- lm(sqrt(finaldata$Happy)~log(finaldata$Household)+bMarital+log(finaldata$Sex)+bJobSat+bIncome+bWorkhrs+sqrt(finaldata$Health))  
om8 <- lm(sqrt(finaldata$Happy)~log(finaldata$Household)+log(finaldata$Instagram)+bMarital+log(finaldata$Sex)+bJobSat+bIncome+bWorkhrs+sqrt(finaldata$Health))  
  
# Adjusted R-square :  
Rad <- rs$adjr2  
  
#Subset size=1  
n <- length(om1$residuals); p <- 1  
#Calculate AIC  
AIC1 <- extractAIC(om1,k=2)[2]  
#Calculate AICc = AIC + 2k(k+1)/(n-k-1)  
AICc1 <- extractAIC(om1,k=2)[2]+2\*(p+2)\*(p+3)/(n-p-1)  
#Calculate BIC  
BIC1 <-extractAIC(om1,k=log(n))[2]  
  
#Subset size=2  
p <-2  
#Calculate AIC  
AIC2 <- extractAIC(om2,k=2)[2]  
#Calculate AICc  
AICc2 <- extractAIC(om2,k=2)[2]+2\*(p+2)\*(p+3)/(n-p-1)  
#Calculate BIC  
BIC2 <- extractAIC(om2,k=log(n))[2]  
  
#Subset size=3  
p <- 3  
#Calculate AIC  
AIC3 <- extractAIC(om3,k=2)[2]  
#Calculate AICc  
AICc3 <- extractAIC(om3,k=2)[2]+2\*(p+2)\*(p+3)/(n-p-1)  
#Calculate BIC  
BIC3 <- extractAIC(om3,k=log(n))[2]  
  
#Subset size=4  
p <- 4  
#Calculate AIC  
AIC4 <- extractAIC(om4,k=2)[2]  
#Calculate AICc  
AICc4 <- extractAIC(om4,k=2)[2]+2\*(p+2)\*(p+3)/(n-p-1)  
#Calculate BIC  
BIC4 <- extractAIC(om4,k=log(n))[2]  
  
#Subset size=5  
p <- 5  
#Calculate AIC  
AIC5 <- extractAIC(om5,k=2)[2]  
#Calculate AICc  
AICc5 <- extractAIC(om5,k=2)[2]+2\*(p+2)\*(p+3)/(n-p-1)  
#Calculate BIC  
BIC5 <- extractAIC(om5,k=log(n))[2]  
  
#Subset size=6  
p <- 6  
#Calculate AIC  
AIC6 <- extractAIC(om6,k=2)[2]  
#Calculate AICc  
AICc6 <- extractAIC(om6,k=2)[2]+2\*(p+2)\*(p+3)/(n-p-1)  
#Calculate BIC  
BIC6 <- extractAIC(om6,k=log(n))[2]  
  
#Subset size=7  
p <- 7  
#Calculate AIC  
AIC7 <- extractAIC(om7,k=2)[2]  
#Calculate AICc  
AICc7 <- extractAIC(om7,k=2)[2]+2\*(p+2)\*(p+3)/(n-p-1)  
#Calculate BIC  
BIC7 <- extractAIC(om7,k=log(n))[2]  
  
#Subset size=8  
p <- 8  
#Calculate AIC  
AIC8 <- extractAIC(om8,k=2)[2]  
#Calculate AICc  
AICc8 <- extractAIC(om8,k=2)[2]+2\*(p+2)\*(p+3)/(n-p-1)  
#Calculate BIC  
BIC8 <- extractAIC(om8,k=log(n))[2]  
  
  
  
AIC <- c(AIC1,AIC2,AIC3,AIC4,AIC5, AIC6, AIC7, AIC8)  
AICc <- c(AICc1,AICc2,AICc3,AICc4,AICc5, AICc6, AICc7, AICc8)  
BIC <- c(BIC1,BIC2,BIC3,BIC4,BIC5, BIC6, BIC7, BIC8)  
  
data.frame(Size=1:8, Radj2=Rad,AIC=AIC, AICc=AICc, BIC=BIC)

## Size Radj2 AIC AICc BIC  
## 1 1 0.06022350 -1143.482 -1143.422 -1135.515  
## 2 2 0.08652574 -1153.758 -1153.657 -1141.807  
## 3 3 0.10553126 -1161.114 -1160.962 -1145.179  
## 4 4 0.11819139 -1165.785 -1165.571 -1145.865  
## 5 5 0.12291154 -1166.930 -1166.643 -1143.026  
## 6 6 0.12424638 -1166.551 -1166.182 -1138.664  
## 7 7 0.12339516 -1165.185 -1164.722 -1133.313  
## 8 8 0.12232015 -1163.720 -1163.153 -1127.865

#Radj2 suggests size 6 and 7 (6 is higher), AIC suggests size 5 and 6 (5 is lower), AICc suggests size 5 and 6 (5 is lower), BIC suggests size 3 and 4 (4 is lower).   
  
#Let's compare the numerical summaries.  
  
summary(om3)

##   
## Call:  
## lm(formula = sqrt(finaldata$Happy) ~ log(finaldata$Household) +   
## bJobSat + sqrt(finaldata$Health))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.50652 -0.24713 0.03021 0.16465 0.49844   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.90785 0.08915 10.183 < 2e-16 \*\*\*  
## log(finaldata$Household) -0.12238 0.02766 -4.424 1.25e-05 \*\*\*  
## bJobSat 0.20025 0.05204 3.848 0.000139 \*\*\*  
## sqrt(finaldata$Health) 0.15829 0.04367 3.625 0.000327 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2305 on 393 degrees of freedom  
## Multiple R-squared: 0.1123, Adjusted R-squared: 0.1055   
## F-statistic: 16.57 on 3 and 393 DF, p-value: 3.691e-10

#all variables and slope significant. adj R^2 is .1034  
summary(om4)

##   
## Call:  
## lm(formula = sqrt(finaldata$Happy) ~ log(finaldata$Household) +   
## bMarital + bJobSat + sqrt(finaldata$Health))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.49783 -0.23311 0.03286 0.14497 0.51917   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.08596 0.11230 9.670 < 2e-16 \*\*\*  
## log(finaldata$Household) -0.08492 0.03107 -2.733 0.006563 \*\*   
## bMarital -0.18380 0.07132 -2.577 0.010323 \*   
## bJobSat 0.18262 0.05212 3.504 0.000511 \*\*\*  
## sqrt(finaldata$Health) 0.13950 0.04397 3.173 0.001629 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2289 on 392 degrees of freedom  
## Multiple R-squared: 0.1271, Adjusted R-squared: 0.1182   
## F-statistic: 14.27 on 4 and 392 DF, p-value: 6.961e-11

##all variables and slope significant. adj R^2 is .1155  
summary(om5)

##   
## Call:  
## lm(formula = sqrt(finaldata$Happy) ~ log(finaldata$Household) +   
## bMarital + log(finaldata$Sex) + bJobSat + sqrt(finaldata$Health))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.51329 -0.22843 0.03429 0.15982 0.54298   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.06168 0.11285 9.408 < 2e-16 \*\*\*  
## log(finaldata$Household) -0.08413 0.03099 -2.715 0.006928 \*\*   
## bMarital -0.18661 0.07114 -2.623 0.009057 \*\*   
## log(finaldata$Sex) 0.05858 0.03322 1.763 0.078614 .   
## bJobSat 0.18023 0.05200 3.466 0.000586 \*\*\*  
## sqrt(finaldata$Health) 0.14525 0.04397 3.303 0.001044 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2283 on 391 degrees of freedom  
## Multiple R-squared: 0.134, Adjusted R-squared: 0.1229   
## F-statistic: 12.1 on 5 and 391 DF, p-value: 6.534e-11

#all variables and slope significant, except Sex. adj R^2 is .1209.  
summary(om6)

##   
## Call:  
## lm(formula = sqrt(finaldata$Happy) ~ log(finaldata$Household) +   
## bMarital + log(finaldata$Sex) + bJobSat + bWorkhrs + sqrt(finaldata$Health))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.50877 -0.22326 0.03641 0.15990 0.54593   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.96150 0.13785 6.975 1.32e-11 \*\*\*  
## log(finaldata$Household) -0.08206 0.03101 -2.646 0.008470 \*\*   
## bMarital -0.18731 0.07109 -2.635 0.008752 \*\*   
## log(finaldata$Sex) 0.05493 0.03332 1.649 0.100047   
## bJobSat 0.17757 0.05200 3.415 0.000705 \*\*\*  
## bWorkhrs 0.24384 0.19301 1.263 0.207230   
## sqrt(finaldata$Health) 0.14501 0.04394 3.300 0.001055 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2281 on 390 degrees of freedom  
## Multiple R-squared: 0.1375, Adjusted R-squared: 0.1242   
## F-statistic: 10.36 on 6 and 390 DF, p-value: 1.153e-10

#all variables and slope significant, except Sex and Workhrs. adj R^2 is .1222.  
summary(om7)

##   
## Call:  
## lm(formula = sqrt(finaldata$Happy) ~ log(finaldata$Household) +   
## bMarital + log(finaldata$Sex) + bJobSat + bIncome + bWorkhrs +   
## sqrt(finaldata$Health))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.52085 -0.22414 0.03811 0.16185 0.53572   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.046436 0.175025 5.979 5.09e-09 \*\*\*  
## log(finaldata$Household) -0.084650 0.031200 -2.713 0.00696 \*\*   
## bMarital -0.175658 0.072646 -2.418 0.01607 \*   
## log(finaldata$Sex) 0.048366 0.034361 1.408 0.16006   
## bJobSat 0.172716 0.052387 3.297 0.00107 \*\*   
## bIncome -0.008879 0.011264 -0.788 0.43105   
## bWorkhrs 0.196926 0.202070 0.975 0.33039   
## sqrt(finaldata$Health) 0.139383 0.044535 3.130 0.00188 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2282 on 389 degrees of freedom  
## Multiple R-squared: 0.1389, Adjusted R-squared: 0.1234   
## F-statistic: 8.963 on 7 and 389 DF, p-value: 2.955e-10

#all variables and slope significant, except Sex, Income, Workhrs. adj R^2 is .1212.  
  
#Stepwise: Forward Selection  
  
mint <- lm(sqrt(finaldata$Happy)~1,data=finaldata)  
forwardAIC <- step(mint,scope=list(lower=~1,   
 upper=~log(finaldata$Household),sqrt(finaldata$Age), log(finaldata$Instagram), bMarital, log(finaldata$Sex), bJobSat, bIncome, bWorkhrs, sqrt(finaldata$Health)), direction="forward", data=finaldata)

## Start: AIC=-1119.82  
## sqrt(finaldata$Happy) ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + log(finaldata$Household) 1 0.97641 22.552 -1134.7  
## <none> 23.528 -1119.8  
##   
## Step: AIC=-1134.65  
## sqrt(finaldata$Happy) ~ log(finaldata$Household)

forwardBIC <- step(mint,scope=list(lower=~1,   
 upper=~log(finaldata$Household),sqrt(finaldata$Age), log(finaldata$Instagram), bMarital, log(finaldata$Sex), bJobSat, bIncome, bWorkhrs, sqrt(finaldata$Health)),  
 direction="forward", data=finaldata,k=log(n))

## Start: AIC=-1115.84  
## sqrt(finaldata$Happy) ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + log(finaldata$Household) 1 0.97641 22.552 -1126.7  
## <none> 23.528 -1115.8  
##   
## Step: AIC=-1126.68  
## sqrt(finaldata$Happy) ~ log(finaldata$Household)

#Stepwise: Backward Selection  
  
backAIC <- step(om8,direction="backward", data=finaldata)

## Start: AIC=-1163.72  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + log(finaldata$Instagram) +   
## bMarital + log(finaldata$Sex) + bJobSat + bIncome + bWorkhrs +   
## sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## - log(finaldata$Instagram) 1 0.02730 20.260 -1165.2  
## - bIncome 1 0.02751 20.260 -1165.2  
## - bWorkhrs 1 0.04540 20.278 -1164.8  
## <none> 20.233 -1163.7  
## - log(finaldata$Sex) 1 0.11376 20.347 -1163.5  
## - bMarital 1 0.31849 20.551 -1159.5  
## - log(finaldata$Household) 1 0.37570 20.609 -1158.4  
## - sqrt(finaldata$Health) 1 0.50423 20.737 -1156.0  
## - bJobSat 1 0.57990 20.813 -1154.5  
##   
## Step: AIC=-1165.18  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
## log(finaldata$Sex) + bJobSat + bIncome + bWorkhrs + sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## - bIncome 1 0.03236 20.293 -1166.5  
## - bWorkhrs 1 0.04946 20.310 -1166.2  
## <none> 20.260 -1165.2  
## - log(finaldata$Sex) 1 0.10319 20.363 -1165.2  
## - bMarital 1 0.30451 20.565 -1161.3  
## - log(finaldata$Household) 1 0.38338 20.644 -1159.7  
## - sqrt(finaldata$Health) 1 0.51017 20.770 -1157.3  
## - bJobSat 1 0.56613 20.826 -1156.2  
##   
## Step: AIC=-1166.55  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
## log(finaldata$Sex) + bJobSat + bWorkhrs + sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## - bWorkhrs 1 0.08304 20.376 -1166.9  
## <none> 20.293 -1166.5  
## - log(finaldata$Sex) 1 0.14141 20.434 -1165.8  
## - bMarital 1 0.36124 20.654 -1161.5  
## - log(finaldata$Household) 1 0.36435 20.657 -1161.5  
## - sqrt(finaldata$Health) 1 0.56672 20.859 -1157.6  
## - bJobSat 1 0.60677 20.899 -1156.8  
##   
## Step: AIC=-1166.93  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
## log(finaldata$Sex) + bJobSat + sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## <none> 20.376 -1166.9  
## - log(finaldata$Sex) 1 0.16205 20.538 -1165.8  
## - bMarital 1 0.35854 20.734 -1162.0  
## - log(finaldata$Household) 1 0.38403 20.760 -1161.5  
## - sqrt(finaldata$Health) 1 0.56862 20.944 -1158.0  
## - bJobSat 1 0.62609 21.002 -1156.9

backBIC <- step(om8,direction="backward", data=finaldata, k=log(n))

## Start: AIC=-1127.86  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + log(finaldata$Instagram) +   
## bMarital + log(finaldata$Sex) + bJobSat + bIncome + bWorkhrs +   
## sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## - log(finaldata$Instagram) 1 0.02730 20.260 -1133.3  
## - bIncome 1 0.02751 20.260 -1133.3  
## - bWorkhrs 1 0.04540 20.278 -1133.0  
## - log(finaldata$Sex) 1 0.11376 20.347 -1131.6  
## <none> 20.233 -1127.9  
## - bMarital 1 0.31849 20.551 -1127.7  
## - log(finaldata$Household) 1 0.37570 20.609 -1126.5  
## - sqrt(finaldata$Health) 1 0.50423 20.737 -1124.1  
## - bJobSat 1 0.57990 20.813 -1122.6  
##   
## Step: AIC=-1133.31  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
## log(finaldata$Sex) + bJobSat + bIncome + bWorkhrs + sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## - bIncome 1 0.03236 20.293 -1138.7  
## - bWorkhrs 1 0.04946 20.310 -1138.3  
## - log(finaldata$Sex) 1 0.10319 20.363 -1137.3  
## - bMarital 1 0.30451 20.565 -1133.4  
## <none> 20.260 -1133.3  
## - log(finaldata$Household) 1 0.38338 20.644 -1131.8  
## - sqrt(finaldata$Health) 1 0.51017 20.770 -1129.4  
## - bJobSat 1 0.56613 20.826 -1128.4  
##   
## Step: AIC=-1138.66  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
## log(finaldata$Sex) + bJobSat + bWorkhrs + sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## - bWorkhrs 1 0.08304 20.376 -1143.0  
## - log(finaldata$Sex) 1 0.14141 20.434 -1141.9  
## <none> 20.293 -1138.7  
## - bMarital 1 0.36124 20.654 -1137.6  
## - log(finaldata$Household) 1 0.36435 20.657 -1137.6  
## - sqrt(finaldata$Health) 1 0.56672 20.859 -1133.7  
## - bJobSat 1 0.60677 20.899 -1133.0  
##   
## Step: AIC=-1143.03  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
## log(finaldata$Sex) + bJobSat + sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## - log(finaldata$Sex) 1 0.16205 20.538 -1145.9  
## <none> 20.376 -1143.0  
## - bMarital 1 0.35854 20.734 -1142.1  
## - log(finaldata$Household) 1 0.38403 20.760 -1141.6  
## - sqrt(finaldata$Health) 1 0.56862 20.944 -1138.1  
## - bJobSat 1 0.62609 21.002 -1137.0  
##   
## Step: AIC=-1145.87  
## sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
## bJobSat + sqrt(finaldata$Health)  
##   
## Df Sum of Sq RSS AIC  
## <none> 20.538 -1145.9  
## - bMarital 1 0.34800 20.886 -1145.2  
## - log(finaldata$Household) 1 0.39129 20.929 -1144.4  
## - sqrt(finaldata$Health) 1 0.52740 21.065 -1141.8  
## - bJobSat 1 0.64325 21.181 -1139.6

#For backward selection, the model below is the best fit.  
sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
 log(finaldata$Sex) + bJobSat + sqrt(finaldata$Health)

## sqrt(finaldata$Happy) ~ log(finaldata$Household) + bMarital +   
## log(finaldata$Sex) + bJobSat + sqrt(finaldata$Health)