#

# IST687: week 8 Synchronous Code

#

# Pulls the data directly into a data frame

juul2 <- read.table("http://people.sc.fsu.edu/~jburkardt/datasets/iswr/juul2.csv",

header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)

# Make sure to double check that the variables are numeric and not factors

str(juul2)

head(juul2)

# This reduces the data set to include only those cases with complete data

# On age, height, igf1, and weight

#

juulComplete <- juul2[complete.cases(juul2[,c(1,4,5,8)]),]

# Makes it easier to use theses data

attach(juulComplete)

# Plot each bivariate combination for predicting igf1

# Make sure to interpret each relationship

plot(igf1 ~ weight, xlab = "Body weight", ylab = "Insulin-like growth factor" )

plot(igf1 ~ height, xlab = "Body height", ylab = "Insulin-like growth factor" )

plot(igf1 ~ age, xlab = "Age", ylab = "Insulin-like growth factor" )

# Run three increasingly complex models

modelW <- lm(igf1 ~ weight)

summary(modelW)

#

#ggplot

plot(igf1 ~ weight, xlab = "Body weight", ylab = "Insulin-like growth factor" )

abline(modelW,col="red",cex=4)

# R-squared .26 and weight does fine as predictor

#

modelHW <- lm(igf1 ~ weight + height)

summary(modelHW)

# R-squared .34 and height takes over from weight

modelHWA <- lm(igf1 ~ weight + height + age)

summary(modelHWA)

# R-squared .41, best model so far

# height and age work together to predict igf1

#

# Parsimonious Analysis - akaike information criterion

#

step(modelHWA,data=juulCom,direction='backward')

#

#

# residual analysis

#

# igf w/ age & height

modelAH<-lm(igf1~age+height)

summary(modelAH)

juulComplete$resid<-residuals(modelAH)

head(juulComplete)

sorted<-juulComplete[order(abs(-juulComplete$resid)),]

head(sorted)

plot(modelW)

plot(modelHW)

plot(modelHWA)