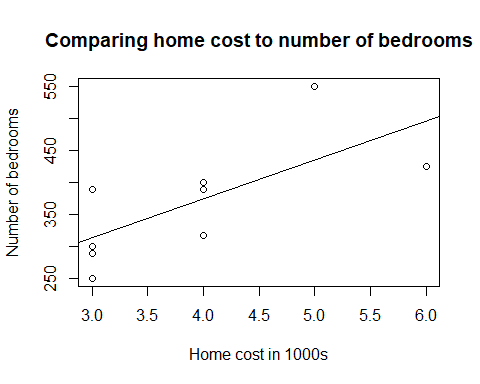
Untitled

2023-03-28

#11.6  
homes <- c(300, 250, 400, 550, 317, 389, 425, 289, 389)  
beds <-c(3, 3, 4, 5, 4, 3, 6, 3, 4)  
homes.reg = lm(homes ~ beds)  
plot(homes~beds, main="Comparing home cost to number of bedrooms", xlab="Home cost in 1000s", ylab="Number of bedrooms")  
abline(homes.reg)



se.h1<-coef(summary(homes.reg))[2,2]  
t.h1=(homes.reg$coefficients[2]-60)/se.h1  
pval.h1= pt(t.h1, df= length(homes)-2)  
cat("Test statistic = ",t.h1, "P-value = ", pval.h1)

## Test statistic = 0.03098394 P-value = 0.5119264

#11.10  
library(UsingR)

## Loading required package: MASS

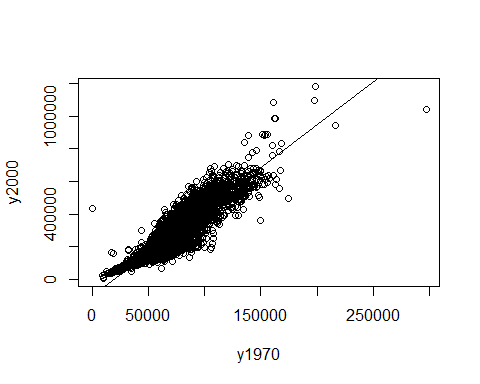
## Loading required package: HistData

## Loading required package: Hmisc

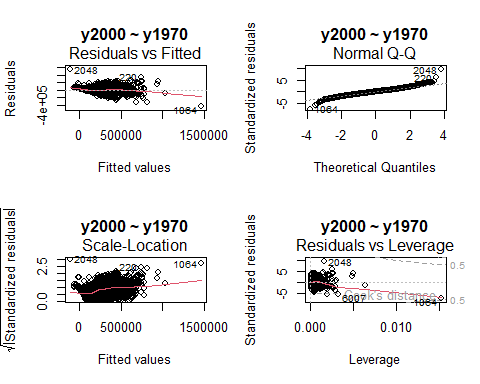
##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, units

attach(homedata)  
hdat.reg = lm(y2000 ~ y1970)  
plot(y2000 ~ y1970, data = homedata)  
abline(hdat.reg)



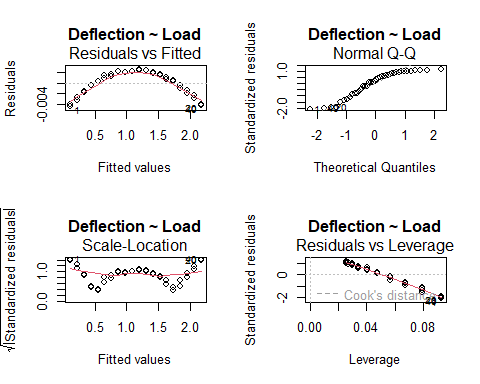
par(mfrow= c(2, 2))  
plot(hdat.reg, main="y2000 ~ y1970")



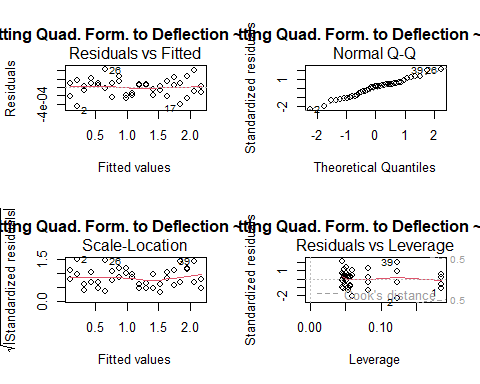
#predict(hdat.reg, myprediction=data.frame(y1970=80000))  
detach(homedata)

11.10 discussion: The linear model appears to be a descent fit. Of the residuals, there looks to be an even amount of random residuals above and below the fitted line. The QQ plot also appears to be fairly straight.

#11.24  
attach(deflection)  
def.reg1= lm(Deflection ~ Load, data = deflection)  
par(mfrow= c(2, 2))  
plot(def.reg1, main="Deflection ~ Load")



def.reg2= lm(Deflection ~ Load + I(Load^2))  
plot(def.reg2, main="Fitting Quad. Form. to Deflection ~ Load")



detach(deflection)

11.24 discussion: The quadratic formula does appear to be a better fit. The residuals are more random above/below the fitted line and the QQ plot is showing a much straighter line.

#4.7  
library(MASS)  
nusacar <- subset(Cars93, Origin !="USA" & Cylinders == 4 & Max.Price <=15, select=c(Make, Origin, Cylinders, Max.Price))  
nusacar

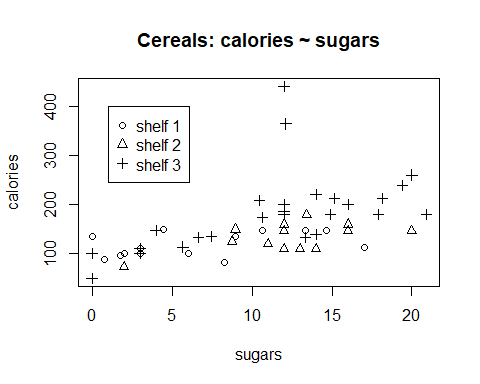
## Make Origin Cylinders Max.Price  
## 40 Geo Storm non-USA 4 13.5  
## 44 Hyundai Excel non-USA 4 9.2  
## 45 Hyundai Elantra non-USA 4 11.0  
## 46 Hyundai Scoupe non-USA 4 11.0  
## 53 Mazda 323 non-USA 4 9.1  
## 54 Mazda Protege non-USA 4 12.3  
## 62 Mitsubishi Mirage non-USA 4 12.9  
## 64 Nissan Sentra non-USA 4 14.9  
## 81 Subaru Loyale non-USA 4 11.3  
## 84 Toyota Tercel non-USA 4 11.8  
## 88 Volkswagen Fox non-USA 4 9.5

#4.13  
library(UsingR)  
attach(batting)  
batavg <- aggregate(H/AB, list(teamID), FUN=mean)  
batavg

## Group.1 x  
## 1 ANA 0.2854710  
## 2 ARI 0.2757978  
## 3 ATL 0.2641674  
## 4 BAL 0.2443030  
## 5 BOS 0.2698326  
## 6 CHA 0.2641604  
## 7 CHN 0.2457606  
## 8 CIN 0.2615390  
## 9 CLE 0.2461690  
## 10 COL 0.2655686  
## 11 DET 0.2436964  
## 12 FLO 0.2658364  
## 13 HOU 0.2669133  
## 14 KCA 0.2520347  
## 15 LAN 0.2744040  
## 16 MIL 0.2639270  
## 17 MIN 0.2647846  
## 18 MON 0.2633824  
## 19 NYA 0.2642199  
## 20 NYN 0.2621652  
## 21 OAK 0.2547011  
## 22 PHI 0.2670113  
## 23 PIT 0.2418561  
## 24 SDN 0.2544134  
## 25 SEA 0.2739472  
## 26 SFN 0.2679086  
## 27 SLN 0.2659466  
## 28 TBA 0.2464189  
## 29 TEX 0.2665969  
## 30 TOR 0.2526405

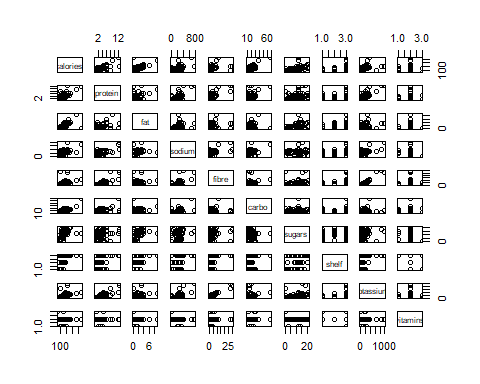
detach(batting)

#5.1  
library(MASS)  
with(UScereal, plot(calories ~ sugars, data = UScereal, pch=shelf, main="Cereals: calories ~ sugars"))  
with(UScereal, legend(1, 400, c("shelf 1", "shelf 2", "shelf 3"), pch=1:3))

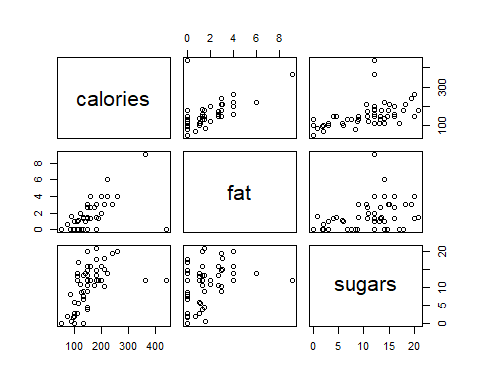


5.1 discussion: There is a definite pattern of sugar affecting calories – as sugar increases, calories increase. In general, it looks like the cereals containing the higher amount of sugar/calories is going to be on shelf 3.

#5.3 this question said to show all numerical values, so I included that, but we really only needed to see calories, fat, and sugar. See the next page for a better graph set.  
library(MASS)  
with(UScereal, pairs(cbind(calories, protein, fat, sodium, fibre, carbo, sugars, shelf, potassium, vitamins)))

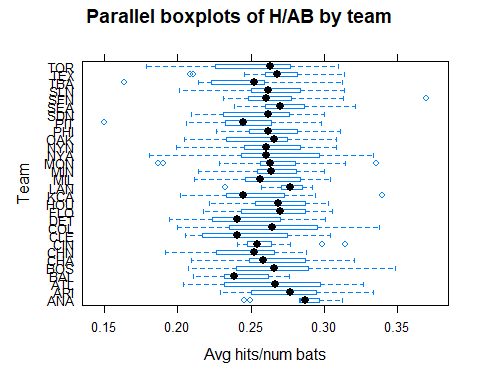


with(UScereal, pairs(cbind(calories, fat, sugars)))



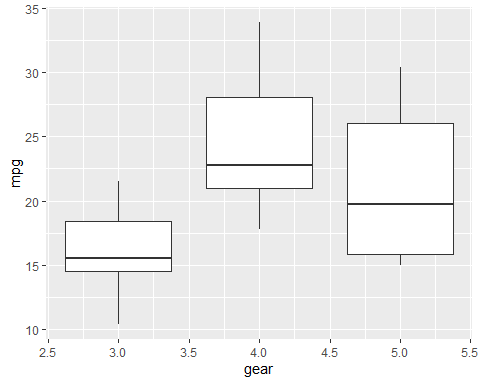
5.3 discussion: The correlation between calories and fat looks strongest. It has fewer outliers, the data has a clear positive direction, and it is in a tidy line.

#5.6  
library(lattice)  
library(UsingR)  
bwplot(factor(teamID) ~ (H/AB), data = batting,main = "Parallel boxplots of H/AB by team", xlab="Avg hits/num bats", ylab="Team")



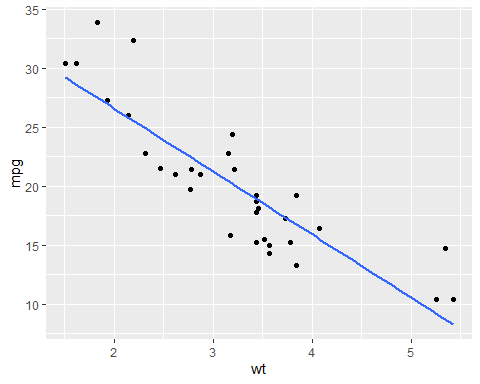
5.6 discussion: The ANA team had the greatest median average.

#5.7.1  
library(ggplot2)  
p <- ggplot(mtcars, aes(x=gear, y=mpg, group=gear)) + geom\_boxplot()  
p



#5.7.2  
library(ggplot2)  
p <- ggplot(mtcars, aes(x=wt, y=mpg )) + geom\_point()  
p + geom\_smooth(method='lm', se=FALSE)

## `geom\_smooth()` using formula = 'y ~ x'



#5.7.3  
library(ggplot2)  
p <- ggplot(mtcars, aes(x=hp, y=mpg )) + geom\_point()  
p + geom\_smooth(method='lm', se=FALSE) +facet\_grid(cyl~gear)

## `geom\_smooth()` using formula = 'y ~ x'

