> bonus <- read.csv(choose.files(), header=TRUE)

> attach(bonus)

> model1 = lm(Value ~ Size + Age + Garage + Rooms + Baths)

rstandard = rstandard(model1)

leverages = hatvalues(model1)

cooks = cooks.distance(model1)

Part a.)

> par(mfrow=c(1,3))

> hist(rstandard)

> hist(leverages)

> hist(cooks)



Part b.)

> rstandard[order(rstandard)]

4 2 5 7 1 9 3 6 8

-1.7087039 -0.3303597 -0.1752872 0.1115119 0.1714062 0.2128637 0.5150996 1.2715418 1.6711654

# There are no outliers as no point exceeds ±2

> leverages[order(leverages)]

2 4 1 8 9 7 3 5 6

0.2982571 0.3881361 0.6172772 0.6318148 0.6869392 0.7282425 0.8063996 0.8578796 0.9850540

# The cutoff point for a high leverage value = 3(k+1)/n = 3(5+1)/9 = 2, no points exceed the cutoff

> cooks[order(cooks)]

7 2 1 9 5 3 4 8 6

0.0055537 0.007731 0.0078976 0.0165707 0.0309114 0.1841938 0.3086821 0.7987505 17.7600361

> qf(.95, 6, 3)

[1] 8.940645

# Observation 6 exceeds the cutoff of 8.94 here for Cook’s Distance

Part c.)

> summary(model1)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 110.1807 154.2452 0.714 0.5266

Size 0.2162 0.0800 2.703 0.0736 .

Age -2.1893 1.0160 -2.155 0.1201

Garage -38.4491 57.7353 -0.666 0.5531

Rooms -0.7231 33.4941 -0.022 0.9841

Baths 1.1132 45.3529 0.025 0.9820

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 49.57 on 3 degrees of freedom

Multiple R-squared: 0.9599, Adjusted R-squared: 0.8931

F-statistic: 14.36 on 5 and 3 DF, p-value: 0.02628

> model1b = lm(Value ~ Size + Age + Garage + Rooms + Baths, subset=-c(6))

> summary(model1b)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 264.6093 163.2638 1.621 0.247

Size -0.2334 0.3015 -0.774 0.520

Age -8.0253 3.9093 -2.053 0.176

Garage 195.8273 160.5691 1.220 0.347

Rooms 123.6370 85.9719 1.438 0.287

Baths -38.9214 45.9141 -0.848 0.486

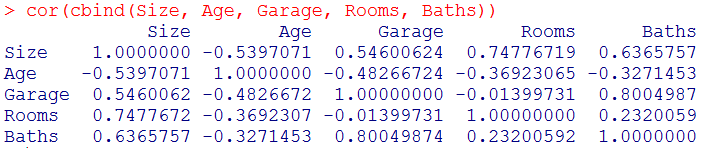
Residual standard error: 41.23 on 2 degrees of freedom

Multiple R-squared: 0.9812, Adjusted R-squared: 0.934

F-statistic: 20.82 on 5 and 2 DF, p-value: 0.04645

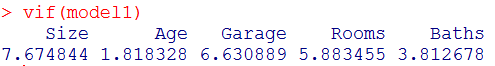
R² has increased from 0.9599 to 0.9812, R²a has increased from 0.8931 to 0.934, s has decreased from 49.57 to 41.23, all of which are good. The F p-value for the model has increased from .02628 to .04645, which is bad. In the original model, if alpha = 0.10, Size was significant. In the 2nd model, no variable is significant.

Part d.)



There are high correlations between Garage/Baths = 0.800 and Size/Rooms = 0.7477 and moderate correlations between Garage/Size =0.546, Baths/Size = 0.6365, and Age/Garage = -.483. Multicollinearity may be a problem here.

Part e.)



Several of the variables have high variance inflation factor values here, but none exceed the cutoff of 10. There is not a case of severe multicollinearity here.