> fit2=lm(medv~lstat+age,data=Boston)

> summary(fit2)

Call:

lm(formula = medv ~ lstat + age, data = Boston)

Residuals:

Min 1Q Median 3Q Max

-15.981 -3.978 -1.283 1.968 23.158

Coefficients:

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

(Intercept) 33.22276 0.73085 45.458 < 2e-16 \*\*\*

lstat -1.03207 0.04819 -21.416 < 2e-16 \*\*\*

age 0.03454 0.01223 2.826 0.00491 \*\*

---

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

Residual standard error: 6.173 on 503 degrees of freedom

Multiple R-squared: 0.5513, Adjusted R-squared: 0.5495

F-statistic: 309 on 2 and 503 DF, p-value: < 2.2e-16

> fit3=lm(medv~.,Boston)

> summary(fit3)

Call:

lm(formula = medv ~ ., data = Boston)

Residuals:

Min 1Q Median 3Q Max

-15.595 -2.730 -0.518 1.777 26.199

Coefficients:

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

(Intercept) 3.646e+01 5.103e+00 7.144 3.28e-12 \*\*\*

crim -1.080e-01 3.286e-02 -3.287 0.001087 \*\*

zn 4.642e-02 1.373e-02 3.382 0.000778 \*\*\*

indus 2.056e-02 6.150e-02 0.334 0.738288

chas 2.687e+00 8.616e-01 3.118 0.001925 \*\*

nox -1.777e+01 3.820e+00 -4.651 4.25e-06 \*\*\*

rm 3.810e+00 4.179e-01 9.116 < 2e-16 \*\*\*

age 6.922e-04 1.321e-02 0.052 0.958229

dis -1.476e+00 1.995e-01 -7.398 6.01e-13 \*\*\*

rad 3.060e-01 6.635e-02 4.613 5.07e-06 \*\*\*

tax -1.233e-02 3.760e-03 -3.280 0.001112 \*\*

ptratio -9.527e-01 1.308e-01 -7.283 1.31e-12 \*\*\*

black 9.312e-03 2.686e-03 3.467 0.000573 \*\*\*

lstat -5.248e-01 5.072e-02 -10.347 < 2e-16 \*\*\*

---

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

Residual standard error: 4.745 on 492 degrees of freedom

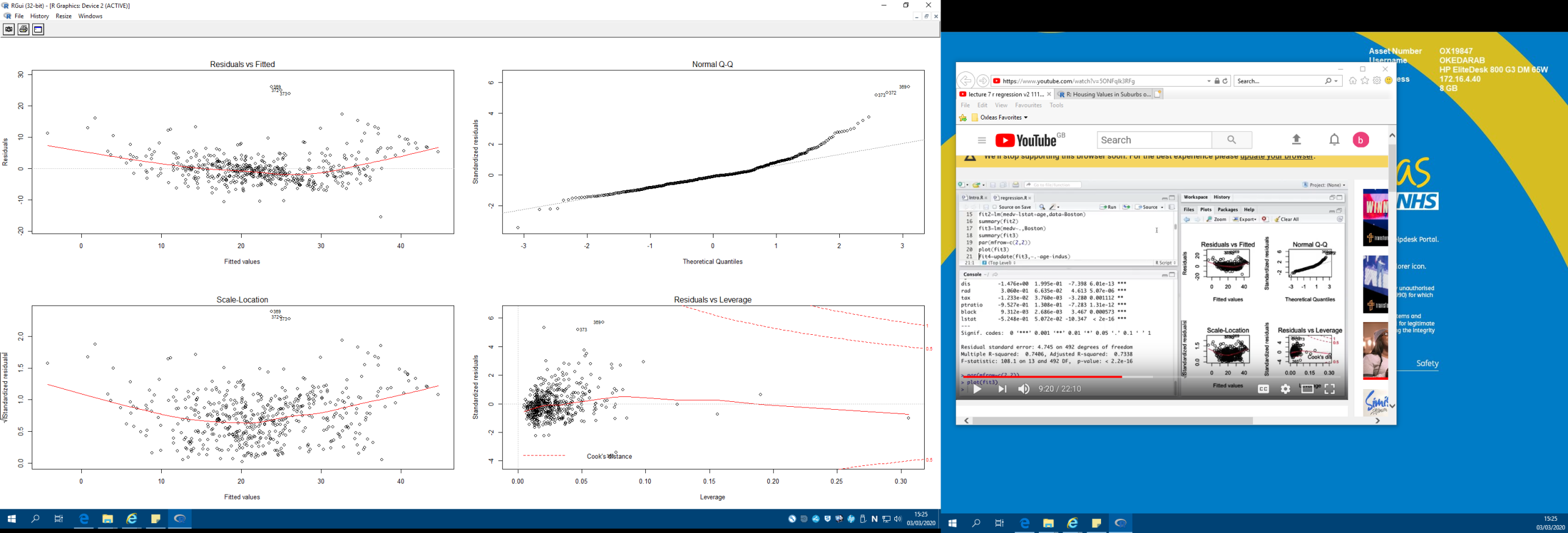
Multiple R-squared: 0.7406, Adjusted R-squared: 0.7338

F-statistic: 108.1 on 13 and 492 DF, p-value: < 2.2e-16

# Note that age is no longer significant (though it was when fitted with lstat and age alone). This suggests that there are a lot of other predictors correlated with age, so in the presence of them, age is no longer required in the model.

> par(mfrow=c(2,2))

> plot(fit3)



> # Note that the first plot suggests that there is some non-linearity in the model

> # Updating the model to remove age and indus

> fit4=update(fit3,~.-age-indus)

> sumaary(fit4)

Error in sumaary(fit4) : could not find function "sumaary"

> summary(fit4)

Call:

lm(formula = medv ~ crim + zn + chas + nox + rm + dis + rad +

tax + ptratio + black + lstat, data = Boston)

Residuals:

Min 1Q Median 3Q Max

-15.5984 -2.7386 -0.5046 1.7273 26.2373

Coefficients:

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

(Intercept) 36.341145 5.067492 7.171 2.73e-12 \*\*\*

crim -0.108413 0.032779 -3.307 0.001010 \*\*

zn 0.045845 0.013523 3.390 0.000754 \*\*\*

chas 2.718716 0.854240 3.183 0.001551 \*\*

nox -17.376023 3.535243 -4.915 1.21e-06 \*\*\*

rm 3.801579 0.406316 9.356 < 2e-16 \*\*\*

dis -1.492711 0.185731 -8.037 6.84e-15 \*\*\*

rad 0.299608 0.063402 4.726 3.00e-06 \*\*\*

tax -0.011778 0.003372 -3.493 0.000521 \*\*\*

ptratio -0.946525 0.129066 -7.334 9.24e-13 \*\*\*

black 0.009291 0.002674 3.475 0.000557 \*\*\*

lstat -0.522553 0.047424 -11.019 < 2e-16 \*\*\*

---

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

Residual standard error: 4.736 on 494 degrees of freedom

Multiple R-squared: 0.7406, Adjusted R-squared: 0.7348

F-statistic: 128.2 on 11 and 494 DF, p-value: < 2.2e-16

> # Note that everything left in the model is now significant

> # Nonlinear terms and interactions

> fit5=lm(medv~lstat\*age,Boston)

> summary(fit5)

Call:

lm(formula = medv ~ lstat \* age, data = Boston)

Residuals:

Min 1Q Median 3Q Max

-15.806 -4.045 -1.333 2.085 27.552

Coefficients:

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

(Intercept) 36.0885359 1.4698355 24.553 < 2e-16 \*\*\*

lstat -1.3921168 0.1674555 -8.313 8.78e-16 \*\*\*

age -0.0007209 0.0198792 -0.036 0.9711

lstat:age 0.0041560 0.0018518 2.244 0.0252 \*

---

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

Residual standard error: 6.149 on 502 degrees of freedom

Multiple R-squared: 0.5557, Adjusted R-squared: 0.5531

F-statistic: 209.3 on 3 and 502 DF, p-value: < 2.2e-16

> fit6=lm(medv~lstat +I(lstat^2),Boston); summary(fit6)

Call:

lm(formula = medv ~ lstat + I(lstat^2), data = Boston)

Residuals:

Min 1Q Median 3Q Max

-15.2834 -3.8313 -0.5295 2.3095 25.4148

Coefficients:

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

(Intercept) 42.862007 0.872084 49.15 <2e-16 \*\*\*

lstat -2.332821 0.123803 -18.84 <2e-16 \*\*\*

I(lstat^2) 0.043547 0.003745 11.63 <2e-16 \*\*\*

---

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

Residual standard error: 5.524 on 503 degrees of freedom

Multiple R-squared: 0.6407, Adjusted R-squared: 0.6393

F-statistic: 448.5 on 2 and 503 DF, p-value: < 2.2e-16

# Note that both the linear and quadratic models are significant

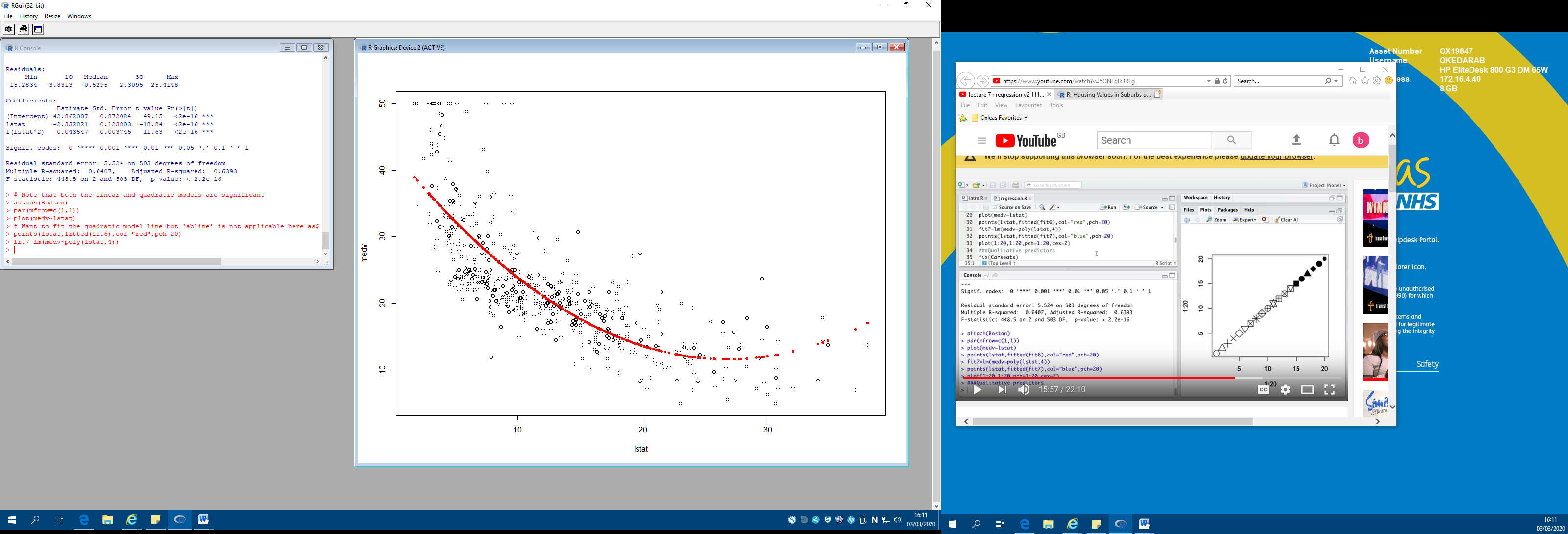
> attach(Boston)

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

> plot(medv~lstat)

# Want to fit the quadratic model line but 'abline' is not applicable here as it only works with straight lines.

> points(lstat,fitted(fit6),col="red",pch=20)



> fit7=lm(medv~poly(lstat,4))

> points(lstat,fitted(fit7),col="blue",pch=20)

