Week 13: Multiple Regression for Home Prices

Data Science 152

Your Name Here

```
library(knitr)
library(dplyr)
library(tidyverse)
library(infer)
library(rstatix)
library(openintro)
library(rgl)
```

Overview

These questions look at a set of data on home sale prices in an area near Durham, North Carolina. It includes information on sales price, lot size, square-footage of living space, number of bedroom and bathrooms, type heating and cooling systems, type of parking and year in which the house was built.

We will try to build a model that predicts price from the other variables in the data set. We'll first focus on using numerical predictors, and then on Thursday add in the categorical variables.

Forward selection model building

The method for building the model follows the "forward selection" procedure outlined in lecture and described in more detail in Chapter 8 of the textbook.

- Start with the predictor variable that gives the largest value of R^2 in a single-predictor regression.
- Try additional variables one at a time, in order from largest single-predictor R^2 to smallest.
- If the added variable increases the value of the adjusted R-squared, R^2_{adj} keep it in the model.

- If the added does not increase R^2_{adj}, do not add it to the model and move on to the next variable.

The single regressions have been done for you, and their results are summarized in **SingleRegressionR2**.

Tuesday: numerical predictors

Exercise 11:

Read in the data from the **DukeForest.csv** file and examine it with `glimpse()` or `summary()`.

```
housing <- read_csv("DukeForest - DukeForest.csv")
```

```
## Rows: 97 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): Heating, Cooling, Parking
## dbl (6): Price1000s, Bed, Bath, AreaSqft, YearBuilt, LotSizeAcre
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
glimpse(housing)
```

```
## Rows: 97
## Columns: 9
## $ Price1000s  <dbl> 1520, 1030, 420, 680, 429, 456, 1270, 557, 698, 650, 540, ~
## $ Bed         <dbl> 3, 5, 2, 4, 4, 3, 5, 4, 4, 3, 4, 4, 3, 5, 4, 5, 3, 4, 4, 3~
## $ Bath        <dbl> 4.0, 4.0, 3.0, 3.0, 3.0, 3.0, 5.0, 3.0, 5.0, 2.0, 3.0, 3.0~
## $ AreaSqft    <dbl> 6040, 4475, 1745, 2091, 1772, 1950, 3909, 2841, 3924, 2173~
## $ YearBuilt   <dbl> 1972, 1969, 1959, 1961, 2020, 2014, 1968, 1973, 1972, 1964~
## $ Heating     <chr> "Gas", "Gas", "Gas", "Other", "Gas", "Gas", "Gas", "Gas", ~
## $ Cooling     <chr> "Central", "Central", "Central", "Central", "Central", "Ce~
## $ Parking     <chr> "Uncovered", "Carport", "Garage", "Carport", "Uncovered", ~
## $ LotSizeAcre <dbl> 0.97, 1.38, 0.51, 0.84, 0.16, 0.45, 0.94, 0.79, 0.53, 0.73~
```

Exercise 22:

Which variables are numerical and which are categorical?

Bed, bath, areasqft, price1000s, and lotacresize are all numerical while heating, cooling, and parking are all categorical.

Exercise 33:

Use the table in **SingleRegressionR2** to determine which of the possible numerical predictors produces the largest value of R^2 in a single-predictor regression. Use `lm()` to run a single-predictor regression with that as the explanatory variable. Use `summary()` to look at output and confirm that you obtain the same value of R^2 as in the table. This will be the starting point for building your multiple regression model.

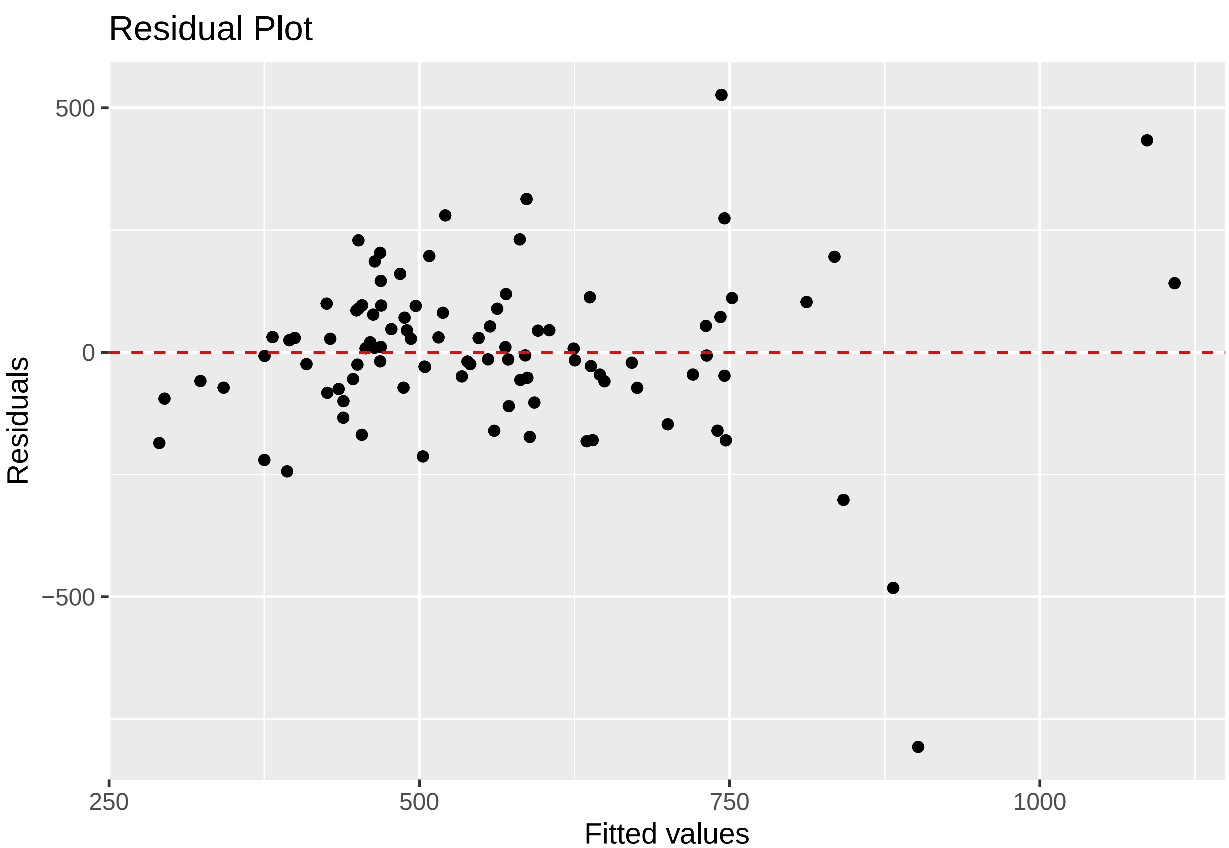
```r
# AreaSqft has the largest R^2 value
housinglm <- lm((Price1000s ~ AreaSqft), housing)
summary(housinglm)
```

```
##
## Call:
## lm(formula = (Price1000s ~ AreaSqft), data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -806.96  -72.29   -6.29   85.68  526.54
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 114.46036   53.34958   2.145   0.0345 *
## AreaSqft      0.16091    0.01823   8.828 5.28e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168.8 on 95 degrees of freedom
## Multiple R-squared:  0.4507, Adjusted R-squared:  0.4449
## F-statistic: 77.93 on 1 and 95 DF,  p-value: 5.278e-14
```

Exercise 44:

Make a residual plot using the `lm` output and the `aes(x = .fitted, y = .resid)` syntax (similar to what you did in the Week 11 classwork). Do you notice any patterns/trends in it?

```
ggplot(housinglm, aes(x = .fitted, y = .resid)) + geom_point() +
    geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
    labs(x = "Fitted values", y = "Residuals", title = "Residual Plot")
```



As the fitted value increases, the likelihood of it being an outlier also increases.

Exercise 55:

Looking at the SingleRegressionR2 table you used in Exercise 3, determine the order in which it would make sense to try the remaining potential numerical predictors.

Bath, then lotacresize, then bed, then yearbuilt.

Exercise 66:

Add your first additional variable so you are doing a multiple regression with two predictors, and examine the output with `summary()`.

```
housinglm2 <- lm(Price1000s ~ AreaSqft + Bath, data = housing)
summary(housinglm2)
```

```
##
```

```
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -806.03  -72.17   -5.58   78.17  513.79
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 36.35826   60.03508   0.606   0.5462
## AreaSqft     0.11910    0.02401   4.960 3.12e-06 ***
## Bath        62.61732   24.28773   2.578   0.0115 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 164 on 94 degrees of freedom
## Multiple R-squared:  0.4869, Adjusted R-squared:  0.476
## F-statistic: 44.61 on 2 and 94 DF,  p-value: 2.389e-14
```

Exercise 77:

Which variable did you add? What value of R^2_{adj} does the 2-predictor model give you? Is it larger than the R^2_{adj} value you got in Exercise 3?

I added Bath, and got an R^2adj of .476 which is slightly larger than what I got in Exercise 3.

Exercise 88:

Add your next variable. If the variable you added in Exercise 6 improved the value of R^2_{adj} keep it so that you are now doing a multiple regression with three predictors.

```r
housinglm3 <- lm(Price1000s ~ AreaSqft + Bath + LotSizeAcre,
    data = housing)
summary(housinglm3)
```

```
##
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -761.71  -80.80    2.00   75.25  480.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -59.83114   60.13058  -0.995 0.322308
## AreaSqft      0.09129    0.02318   3.939 0.000158 ***
## Bath         59.70164   22.44844   2.660 0.009216 **
## LotSizeAcre 319.19704   77.09323   4.140 7.62e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 151.5 on 93 degrees of freedom
```

```
## Multiple R-squared:  0.5668, Adjusted R-squared:  0.5528
## F-statistic: 40.56 on 3 and 93 DF,  p-value: < 2.2e-16
```

Exercise 99:

Which variable did you add? What value of R^2_{adj} does this model give you? Is it larger than your previous highest R^2_{adj}?

I added LotSizeAcre and got .5528, which is larger than my previous R^2adj value.

Exercise 1010:

Add your next variable, keeping your most recent variable if it improved the value of R^2_{adj}, but dropping it if not.

```r
housinglm4 <- lm(Price1000s ~ AreaSqft + Bath + LotSizeAcre +
    Bed, data = housing)
summary(housinglm4)
```

```
##
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre + Bed,
##     data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -771.17  -84.44    7.93   72.50  456.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -32.08876   82.48646  -0.389  0.69816
## AreaSqft      0.09428    0.02405   3.921  0.00017 ***
## Bath         64.43017   24.49139   2.631  0.00999 **
## LotSizeAcre 318.34716   77.42778   4.112 8.53e-05 ***
## Bed         -13.41194   27.17071  -0.494  0.62275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 152.1 on 92 degrees of freedom
## Multiple R-squared:  0.5679, Adjusted R-squared:  0.5491
## F-statistic: 30.23 on 4 and 92 DF,  p-value: 4.664e-16
```

Exercise 1111:

Which variable did you add? What value of R^2_{adj} does this model give you? Is it larger than the than your previous highest R^2_{adj}?

I got .5491 which is slightly lower than my previous R^2adj value.

Exercise 1212:

Add your final variable, keeping your most recent variable if it improved the value of R^2_{adj}, but dropping it if not.

```
housinglm5 <- lm(Price1000s ~ AreaSqft + Bath + LotSizeAcre +
    YearBuilt, data = housing)
summary(housinglm3)
```

```
##
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -761.71  -80.80    2.00   75.25  480.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -59.83114   60.13058  -0.995 0.322308
## AreaSqft      0.09129    0.02318   3.939 0.000158 ***
## Bath         59.70164   22.44844   2.660 0.009216 **
## LotSizeAcre 319.19704   77.09323   4.140 7.62e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 151.5 on 93 degrees of freedom
## Multiple R-squared:  0.5668, Adjusted R-squared:  0.5528
## F-statistic: 40.56 on 3 and 93 DF,  p-value: < 2.2e-16
```

Exercise 1313:

Which variable did you add? What value of R^2_{adj} does this model give you? Is it larger than the than your previous highest R^2_{adj}?

I added YearBuilt and got a higehr R^2adj value of .5631.

Exercise 1414:

If you dropped any variables because they didn't improved R^2_{adj}, try adding them (one at a time, if more than one) to your current best model, and keep them if they improve R^2_{adj}.

```
housinglm6 <- lm(Price1000s ~ AreaSqft + Bath + LotSizeAcre +
    YearBuilt + Bed, data = housing)
summary(housinglm6)
```

```
##
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre + YearBuilt +
##     Bed, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -760.92  -71.14   -4.83   68.10  446.86
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.305e+03  1.847e+03  -1.790  0.07678 .
## AreaSqft     9.408e-02  2.377e-02   3.958  0.00015 ***
## Bath         5.179e+01  2.524e+01   2.052  0.04303 *
## LotSizeAcre  3.505e+02  7.865e+01   4.456 2.37e-05 ***
## YearBuilt    1.674e+00  9.435e-01   1.774  0.07935 .
## Bed         -1.277e+01  2.686e+01  -0.475  0.63559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 150.4 on 91 degrees of freedom
## Multiple R-squared:  0.5824, Adjusted R-squared:  0.5594
## F-statistic: 25.38 on 5 and 91 DF,  p-value: 6.086e-16
```

Exercise 1515:

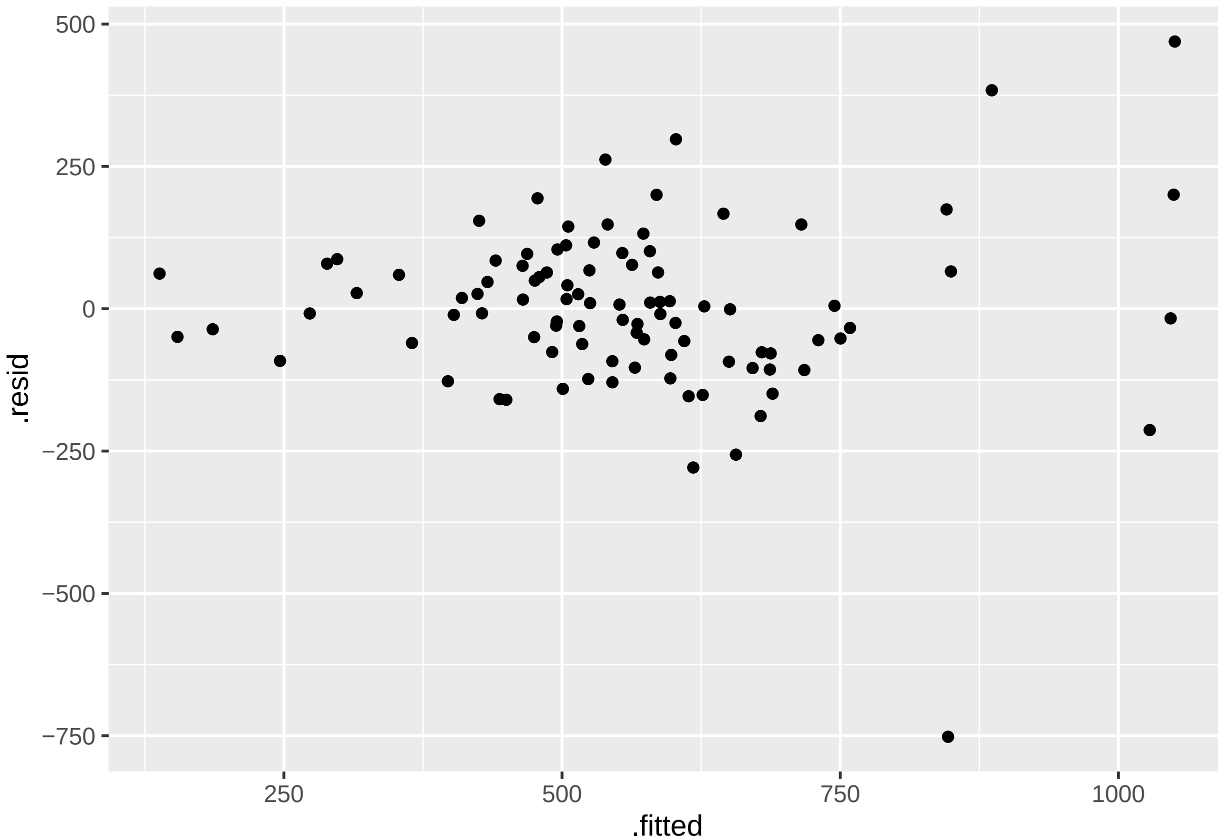
What was the largest value of R^2_{adj} that you obtained in any of these tests? What predictor variables are included in the model that produces the largest R^2_{adj}?

My largest R^2adj value was .5631 when I used Bath, YearBuilt, LotSizeAcre, and AreaSqft

Exercise 1616:

Make a residual plot for this model. How does it differ from the one you made in Exercise 4?

```r
ggplot(housinglm5, aes(x = .fitted, y = .resid)) + geom_point()
```



It is similar but it is much more condensed and there are fewer outliers.

Exercise 1717:

Use this model to predict the price of a 2000 square foot house on a 0.35 acre lot built in 1961 with 3 bedrooms and 2.5 baths.

```
sqftprice <- (0.09124 * 2000)
sizeprice <- (351.4 * 0.35)
bathprice <- (47.24 * 2.5)
yearprice <- (1.68 * 1961)
houseprice <- sqftprice + sizeprice + bathprice + yearprice +
    -3305
print(houseprice)
```

```
## [1] 413.05
```

Total price of $413,050

Thursday: numerical predictors

Exercise 1818:

Add the categorical variable Cooling to the best model you found on Tuesday. Does it produce an improved R^2_{adj} value?

```
housinglm7 <- lm(Price1000s ~ AreaSqft + Bath + LotSizeAcre +
    YearBuilt + Bed + Cooling, data = housing)
summary(housinglm7)
```

```
##
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre + YearBuilt +
##     Bed + Cooling, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -723.35  -77.99   -1.22   85.45  383.42
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.825e+03  1.791e+03  -1.578  0.11817
## AreaSqft      1.022e-01  2.313e-02   4.416 2.79e-05 ***
## Bath          4.108e+01  2.466e+01   1.666  0.09923 .
## LotSizeAcre   3.560e+02  7.594e+01   4.687 9.80e-06 ***
## YearBuilt     1.458e+00  9.140e-01   1.596  0.11406
## Bed          -1.371e+01  2.593e+01  -0.529  0.59835
## CoolingOther -8.406e+01  3.034e+01  -2.771  0.00679 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 145.1 on 90 degrees of freedom
## Multiple R-squared:  0.6152, Adjusted R-squared:  0.5896
## F-statistic: 23.98 on 6 and 90 DF,  p-value: < 2.2e-16
```

Yes it does, the R^2 value is .5896, higher than we've seen before.3

Exercise 1919:

The levels of the Cooling variable are "Central" and "Other." Hopefully you found that the coefficient `CoolingOther` was negative. What does this indicate?

This indicates that the house doesn't have a reliable cooling system built in, making the house less expensive. It also increases the chance of outliers, since there more unknowns regarding it, like what cooling system it does have, if it works reliably, etc. If the house does have cooling, multiplying it by 0 removes the negative value, but if it does not having cooling, multiplying it by 1 makes sure that the negative value does affect the final price.

Exercise 2020:

A) Use your best-fit model to predict the price of a 2000 square foot house on a 0.35 acre lot built in 1961 with 3 bedrooms and 2.5 baths with central air conditioning.

```r
sqftprice <- (0.09124 * 2000)
sizeprice <- (351.4 * 0.35)
bathprice <- (47.24 * 2.5)
yearprice <- (1.68 * 1961)
coolingprice <- (-84.06 * 0)
finalprice2 <- sqftprice + sizeprice + bathprice + yearprice +
    coolingprice + -3305
print(finalprice2)
```

```
## [1] 413.05
```

$413,050.00

B) Use your best-fit model to predict the price of a 2000 square foot house on a 0.35 acre lot built in 1961 with 3 bedrooms and 2.5 baths that does not have central air conditioning (ie has Cooling=Other).

```r
sqftprice <- (0.09124 * 2000)
sizeprice <- (351.4 * 0.35)
bathprice <- (47.24 * 2.5)
yearprice <- (1.68 * 1961)
coolingprice <- (-84.06 * 1)
finalprice3 <- sqftprice + sizeprice + bathprice + yearprice +
    coolingprice + -3305
print(finalprice3)
```

```
## [1] 328.99
```

$328, 990.00

Exercise 2121:

Add the categorical variable Heating to the model (keeping the variable Cooling in the model if it improved the value of R^2_{adj}). Does adding Heating produce an improved R^2_{adj} value?

```r
housinglm8 <- lm(Price1000s ~ AreaSqft + Bath + LotSizeAcre +
    YearBuilt + Bed + Cooling + Heating, data = housing)
summary(housinglm8)
```

```
##
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre + YearBuilt +
##     Bed + Cooling + Heating, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -658.88  -72.89   -0.08   68.90  381.12
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -2.260e+03  1.839e+03  -1.229   0.2224
## AreaSqft         1.058e-01  2.338e-02   4.526 1.89e-05 ***
## Bath             3.392e+01  2.535e+01   1.338   0.1843
## LotSizeAcre      3.367e+02  7.740e+01   4.351 3.68e-05 ***
## YearBuilt        1.211e+00  9.383e-01   1.291   0.2000
## Bed             -1.455e+01  2.599e+01  -0.560   0.5769
## CoolingOther    -7.268e+01  3.243e+01  -2.241   0.0275 *
## HeatingGas      -4.814e+01  6.102e+01  -0.789   0.4322
## HeatingMultiple -5.671e+01  6.194e+01  -0.916   0.3624
## HeatingOther    -1.247e+02  6.849e+01  -1.821   0.0721 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 144.4 on 87 degrees of freedom
## Multiple R-squared:  0.6318, Adjusted R-squared:  0.5937
## F-statistic: 16.58 on 9 and 87 DF,  p-value: 1.564e-15
```

Yes, the R^2 value is .5937, which is slightly higher than the test without Heating.

Exercise 2222:

Add the categorical variable Parking to the model (keeping the variables Cooling and Heating in the model if they improved the value of R^2_{adj}). Does adding Parking produce an improved R^2_{adj} value?

```r
housinglm9 <- lm(Price1000s ~ AreaSqft + Bath + LotSizeAcre +
    YearBuilt + Bed + Cooling + Heating + Parking, data = housing)
summary(housinglm9)
```

```
##
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre + YearBuilt +
##     Bed + Cooling + Heating + Parking, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -660.13  -68.30   -3.38   63.08  373.73
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.183e+03  1.871e+03  -1.167   0.2466
## AreaSqft          1.060e-01  2.365e-02   4.481 2.30e-05 ***
## Bath              3.297e+01  2.581e+01   1.277   0.2051
## LotSizeAcre       3.414e+02  7.879e+01   4.333 4.01e-05 ***
## YearBuilt         1.177e+00  9.524e-01   1.236   0.2198
## Bed              -1.433e+01  2.634e+01  -0.544   0.5877
## CoolingOther     -7.816e+01  3.491e+01  -2.239   0.0278 *
## HeatingGas       -5.439e+01  6.353e+01  -0.856   0.3944
## HeatingMultiple  -6.135e+01  6.331e+01  -0.969   0.3353
## HeatingOther     -1.342e+02  7.185e+01  -1.868   0.0652 .
## ParkingGarage    -1.179e+01  4.558e+01  -0.259   0.7965
## ParkingUncovered  6.306e+00  4.605e+01   0.137   0.8914
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 145.9 on 85 degrees of freedom
## Multiple R-squared:  0.6328, Adjusted R-squared:  0.5853
## F-statistic: 13.32 on 11 and 85 DF,  p-value: 2.343e-14
```

It actually gets smaller with the Parking variable.

Exercise 2323:

What explanatory variables are included in the model that gives the largest value of R^2_{adj}. Run `lm()` one more with this model and make a residual plot.

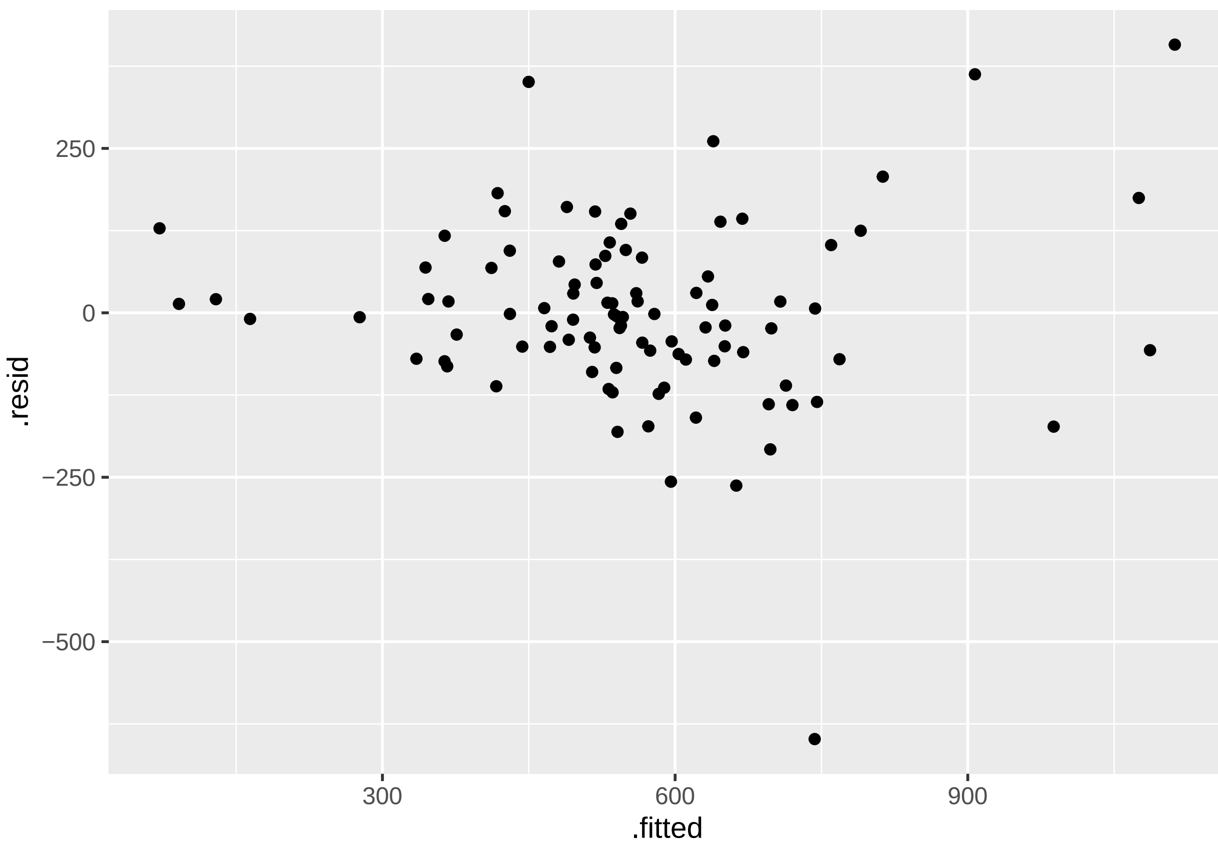
AreaSqft, Bath, LotSizeAcre, YearBuilt, Cooling,and Heating all create the argest R^2 value.

```r
housinglm10 <- lm(Price1000s ~ AreaSqft + Bath + LotSizeAcre +
    YearBuilt + Cooling + Heating, data = housing)
summary(housinglm10)
```

```
##
## Call:
```

```
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre + YearBuilt +
##     Cooling + Heating, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -648.07  -69.83   -6.41   73.52  407.76
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -2.315e+03  1.829e+03  -1.266   0.2089
## AreaSqft         1.025e-01  2.254e-02   4.549 1.72e-05 ***
## Bath             2.863e+01  2.343e+01   1.222   0.2250
## LotSizeAcre      3.384e+02  7.704e+01   4.392 3.12e-05 ***
## YearBuilt        1.223e+00  9.343e-01   1.309   0.1938
## CoolingOther    -7.195e+01  3.227e+01  -2.229   0.0283 *
## HeatingGas      -4.721e+01  6.076e+01  -0.777   0.4392
## HeatingMultiple -5.372e+01  6.147e+01  -0.874   0.3845
## HeatingOther    -1.238e+02  6.821e+01  -1.815   0.0730 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 143.8 on 88 degrees of freedom
## Multiple R-squared:  0.6304, Adjusted R-squared:  0.5968
## F-statistic: 18.76 on 8 and 88 DF,  p-value: 4.022e-16
```

```
ggplot(housinglm10, aes(x = .fitted, y = .resid)) + geom_point()
```



Exercise 2424:

Do you notice any patterns or trends in the residual plot? How does it compare to the plot you made in Exercise 16?

Its slightly more condensed, and there are far fewer outliers than the plot from Exercise 16.

Exercise 2525:

Use your best-fit model to predict the price of a 2000 square foot house on a 0.35 acre lot built in 1961 with 3 bedrooms and 2.5 baths, central air conditioning, gas heating and a garage.

```
summary(housinglm10)
```

```
##
## Call:
## lm(formula = Price1000s ~ AreaSqft + Bath + LotSizeAcre + YearBuilt +
##     Cooling + Heating, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -648.07  -69.83   -6.41   73.52  407.76
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.315e+03  1.829e+03  -1.266   0.2089
## AreaSqft          1.025e-01  2.254e-02   4.549 1.72e-05 ***
## Bath              2.863e+01  2.343e+01   1.222   0.2250
## LotSizeAcre       3.384e+02  7.704e+01   4.392 3.12e-05 ***
## YearBuilt         1.223e+00  9.343e-01   1.309   0.1938
## CoolingOther     -7.195e+01  3.227e+01  -2.229   0.0283 *
## HeatingGas       -4.721e+01  6.076e+01  -0.777   0.4392
## HeatingMultiple -5.372e+01  6.147e+01  -0.874   0.3845
## HeatingOther     -1.238e+02  6.821e+01  -1.815   0.0730 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 143.8 on 88 degrees of freedom
## Multiple R-squared:  0.6304, Adjusted R-squared:  0.5968
## F-statistic: 18.76 on 8 and 88 DF,  p-value: 4.022e-16
```

```
sqftprice <- (0.106 * 2000)
sizeprice <- (341.4 * 0.35)
bathprice <- (32.97 * 2.5)
yearprice <- (1.177 * 1961)
coolingprice <- (-78.16 * 0)
heatingprice <- (-54.39)
finalprice3 <- sqftprice + sizeprice + bathprice + yearprice +
    coolingprice + heatingprice + -2183
```

$484,622.00