

SSC 442 Lab 02

Team 7

1/30/2020

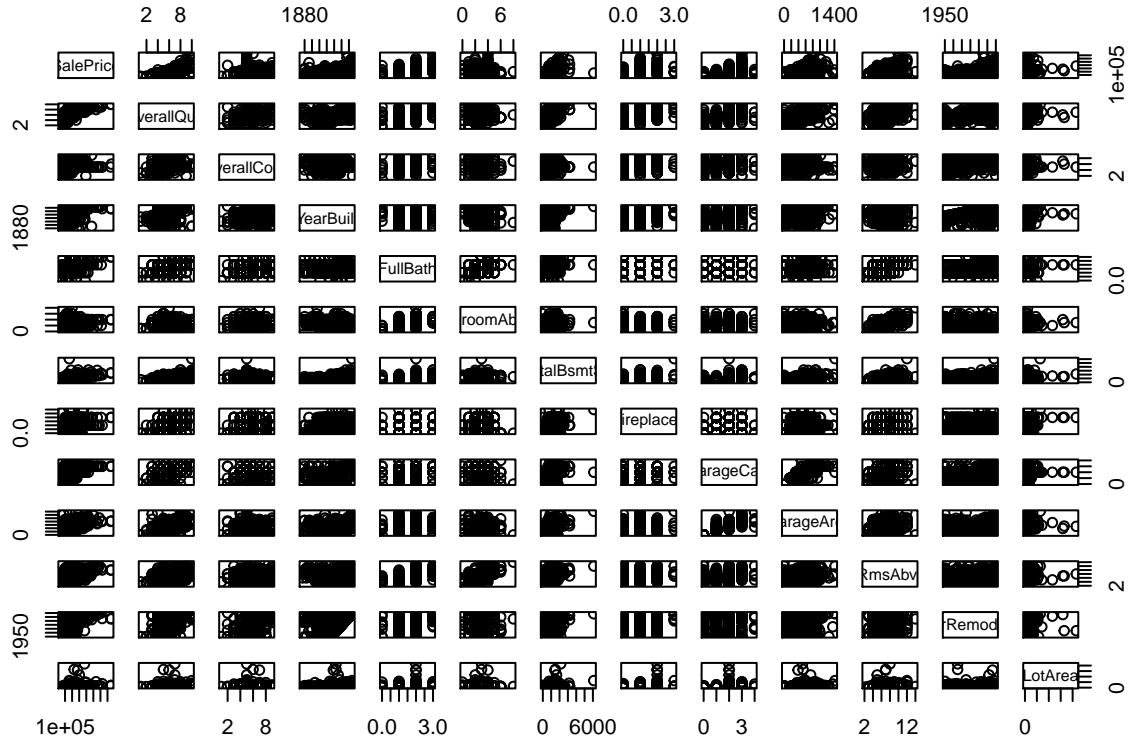
Exercise 1

Part 1:

```
## [1] "MSSubClass"      "LotFrontage"     "LotArea"        "OverallQual"
## [5] "OverallCond"     "YearBuilt"       "YearRemodAdd"   "BsmtFinSF1"
## [9] "BsmtFinSF2"       "BsmtUnfSF"       "TotalBsmtSF"    "X1stFlrSF"
## [13] "X2ndFlrSF"        "LowQualFinSF"    "GrLivArea"      "BsmtFullBath"
## [17] "BsmtHalfBath"     "FullBath"        "HalfBath"       "BedroomAbvGr"
## [21] "KitchenAbvGr"    "TotRmsAbvGrd"   "Fireplaces"     "GarageYrBlt"
## [25] "GarageCars"       "GarageArea"      "WoodDeckSF"     "OpenPorchSF"
## [29] "EnclosedPorch"    "X3SsnPorch"     "ScreenPorch"    "PoolArea"
## [33] "MoSold"           "YrSold"          "SalePrice"
```

The values can be negative; for example, having a bedroom above the garage would be associated with a 4% decrease in sale price rather than causing the value to go up.

Part 2:



Part 3:

```
##          LotArea OverallQual OverallCond YearBuilt FullBath
## LotArea      1.00000000  0.10580574 -0.00563627  0.01422765  0.1260306
## OverallQual   0.10580574  1.00000000 -0.09193234  0.57232277  0.5505997
```

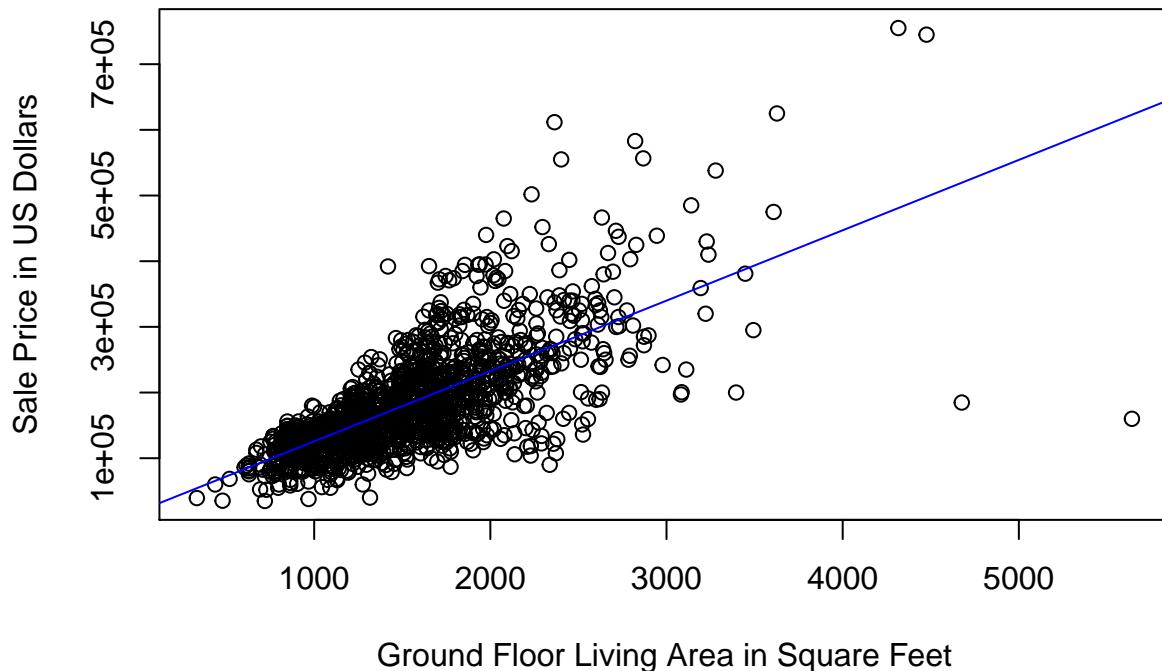
```

## OverallCond -0.00563627 -0.09193234 1.00000000 -0.37598320 -0.1941495
## YearBuilt 0.01422765 0.57232277 -0.37598320 1.00000000 0.4682708
## FullBath 0.12603063 0.55059971 -0.19414949 0.46827079 1.0000000
## BedroomAbvGr 0.11968991 0.10167636 0.01298006 -0.07065122 0.3632520
## TotalBsmtSF 0.26083313 0.53780850 -0.17109751 0.39145200 0.3237224
## Fireplaces 0.27136401 0.39676504 -0.02381998 0.14771640 0.2436705
## GarageCars 0.15487074 0.60067072 -0.18575751 0.53785009 0.4696720
## GarageArea 0.18040276 0.56202176 -0.15152137 0.47895382 0.4056562
## TotRmsAbvGrd 0.19001478 0.42745234 -0.05758317 0.09558913 0.5547843
## YearRemodAdd 0.01378843 0.55068392 0.07374150 0.59285498 0.4390465
## BedroomAbvGr TotalBsmtSF Fireplaces GarageCars GarageArea
## LotArea 0.11968991 0.26083313 0.27136401 0.15487074 0.18040276
## OverallQual 0.10167636 0.53780850 0.39676504 0.60067072 0.56202176
## OverallCond 0.01298006 -0.17109751 -0.02381998 -0.18575751 -0.15152137
## YearBuilt -0.07065122 0.39145200 0.14771640 0.53785009 0.47895382
## FullBath 0.36325198 0.32372241 0.24367050 0.46967204 0.40565621
## BedroomAbvGr 1.00000000 0.05044996 0.10756968 0.08610644 0.06525253
## TotalBsmtSF 0.05044996 1.00000000 0.33951932 0.43458483 0.48666546
## Fireplaces 0.10756968 0.33951932 1.00000000 0.30078877 0.26914124
## GarageCars 0.08610644 0.43458483 0.30078877 1.00000000 0.88247541
## GarageArea 0.06525253 0.48666546 0.26914124 0.88247541 1.00000000
## TotRmsAbvGrd 0.67661994 0.28557256 0.32611448 0.36228857 0.33782212
## YearRemodAdd -0.04058093 0.29106558 0.11258132 0.42062215 0.37159981
## TotRmsAbvGrd YearRemodAdd
## LotArea 0.19001478 0.01378843
## OverallQual 0.42745234 0.55068392
## OverallCond -0.05758317 0.07374150
## YearBuilt 0.09558913 0.59285498
## FullBath 0.55478425 0.43904648
## BedroomAbvGr 0.67661994 -0.04058093
## TotalBsmtSF 0.28557256 0.29106558
## Fireplaces 0.32611448 0.11258132
## GarageCars 0.36228857 0.42062215
## GarageArea 0.33782212 0.37159981
## TotRmsAbvGrd 1.00000000 0.19173982
## YearRemodAdd 0.19173982 1.00000000

```

The correlations match our beliefs; variables such as Year Built and Overall Quality have higher correlation values (.55 and .59). This means that 55% and 59% of the variation in sale price can be explained by having better quality houses and houses built more recently.

Part 4:



The scatterplot suggests a positive correlation between sale price and ground floor living area. The largest outlier above the regression line was a house valued at over \$700,000. This is most likely explained by a combination of ground floor area along with different variables such as overall quality.

Exercise 2

Part 1:

```
##  
## Call:  
## lm(formula = SalePrice ~ GarageInside - 1, data = ameslist)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -150409  -35402   30598  118000  548598  
##  
## Coefficients:  
##                 Estimate Std. Error t value Pr(>|t|)  
## GarageInside    206402     3247    63.57 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 101800 on 1459 degrees of freedom  
## Multiple R-squared:  0.7347, Adjusted R-squared:  0.7345  
## F-statistic:  4041 on 1 and 1459 DF, p-value: < 2.2e-16
```

Part 2:

```
##  
## Call:  
## lm(formula = SalePrice ~ ., data = Ames)
```

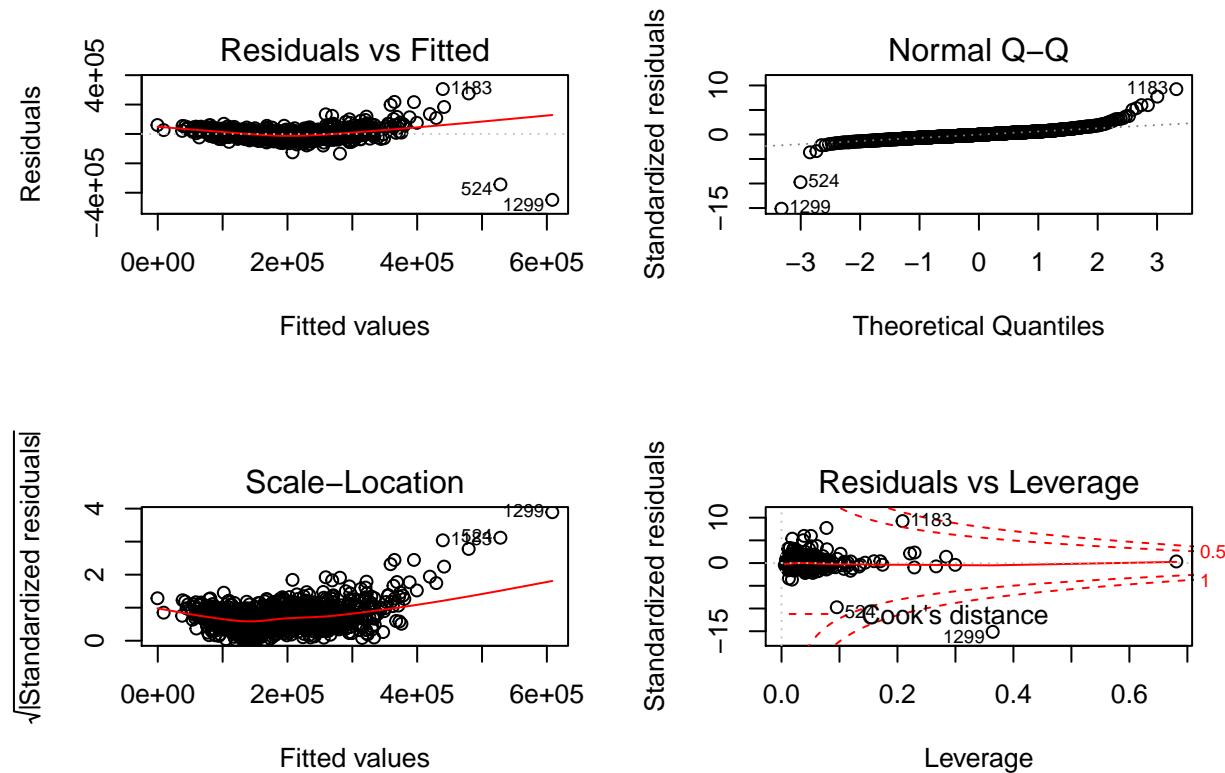
```

##
## Residuals:
##      Min       1Q   Median      3Q      Max
## -448366  -17235   -2368  14868  305479
##
## Coefficients: (2 not defined because of singularities)
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.789e+03  1.713e+06   0.005 0.996373
## MSSubClass    -1.897e+02  3.466e+01  -5.473 5.49e-08 ***
## LotFrontage   -1.090e+02  6.174e+01  -1.765 0.077788 .
## LotArea        5.018e-01  1.586e-01   3.163 0.001603 **
## OverallQual   1.943e+04  1.483e+03  13.108 < 2e-16 ***
## OverallCond   5.194e+03  1.375e+03   3.777 0.000168 ***
## YearBuilt     3.585e+02  8.766e+01   4.089 4.64e-05 ***
## YearRemodAdd  8.512e+01  8.712e+01   0.977 0.328750
## BsmtFinSF1   2.080e+01  5.846e+00   3.558 0.000389 ***
## BsmtFinSF2   8.635e+00  8.845e+00   0.976 0.329181
## BsmtUnfSF    6.263e+00  5.313e+00   1.179 0.238736
## TotalBsmtSF      NA       NA       NA       NA
## X1stFlrSF     4.937e+01  7.380e+00   6.690 3.55e-11 ***
## X2ndFlrSF     4.904e+01  6.104e+00   8.035 2.41e-15 ***
## LowQualFinSF  3.071e+01  2.812e+01   1.092 0.275070
## GrLivArea       NA       NA       NA       NA
## BsmtFullBath  7.852e+03  3.214e+03   2.443 0.014704 *
## BsmtHalfBath  2.173e+03  5.120e+03   0.424 0.671332
## FullBath       4.581e+03  3.551e+03   1.290 0.197236
## HalfBath      -1.451e+02  3.329e+03  -0.044 0.965249
## BedroomAbvGr -1.048e+04  2.174e+03  -4.824 1.61e-06 ***
## KitchenAbvGr -2.248e+04  6.766e+03  -3.322 0.000924 ***
## TotRmsAbvGrd  5.496e+03  1.496e+03   3.674 0.000250 ***
## Fireplaces    4.180e+03  2.199e+03   1.901 0.057561 .
## GarageYrBlt   -7.611e+01  9.147e+01  -0.832 0.405508
## GarageCars    1.717e+04  3.517e+03   4.882 1.21e-06 ***
## GarageArea    1.072e+01  1.219e+01   0.880 0.379163
## WoodDeckSF    2.296e+01  1.009e+01   2.276 0.023036 *
## OpenPorchSF   -8.400e+00  1.947e+01  -0.432 0.666164
## EnclosedPorch 5.793e+00  2.081e+01   0.278 0.780771
## X3SsnPorch    3.561e+01  3.781e+01   0.942 0.346474
## ScreenPorch   5.794e+01  2.033e+01   2.850 0.004453 **
## PoolArea      -7.590e+01  2.974e+01  -2.552 0.010850 *
## MoSold        -2.310e+02  4.245e+02  -0.544 0.586393
## YrSold        -4.021e+02  8.520e+02  -0.472 0.637073
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37150 on 1094 degrees of freedom
##   (333 observations deleted due to missingness)
## Multiple R-squared:  0.8063, Adjusted R-squared:  0.8007
## F-statistic: 142.4 on 32 and 1094 DF,  p-value: < 2.2e-16

```

Some predictors have stronger relationships with sale price. These values add more or less value to sale price based on whether or not they are included in a given home. Overall quality, full bath, year built, and garage cars have statistically significant relationships with the response. The coefficient for year built is 0.59. This means that 59% of the variation in sale price can be attributed to the year that the house was built.

Part 3:



The fit contains too many variables to make an accurate prediction of the average sale price based on ground level living area. Additionally, some outliers are far above and below the fitted regression line. This suggests that some houses sold for much more or much less than the predicted value for a house of that size. The plot does not account for changes in sale price based on seclusion or population density within the area.

Part 4:

Lot Area, Overall Quality, Overall Condition, Basement Full Bath, Kitchen Above Ground, Total Rooms Above Ground, Garage Cars, Screen Porch, and Pool Area are all statistically significant.

Part 5:

Lot area becomes more significant when the $\log(x)$ is taken of the variable. Overall condition becomes less significant when it is squared. Bedroom above ground becomes less significant when the square root is taken.