

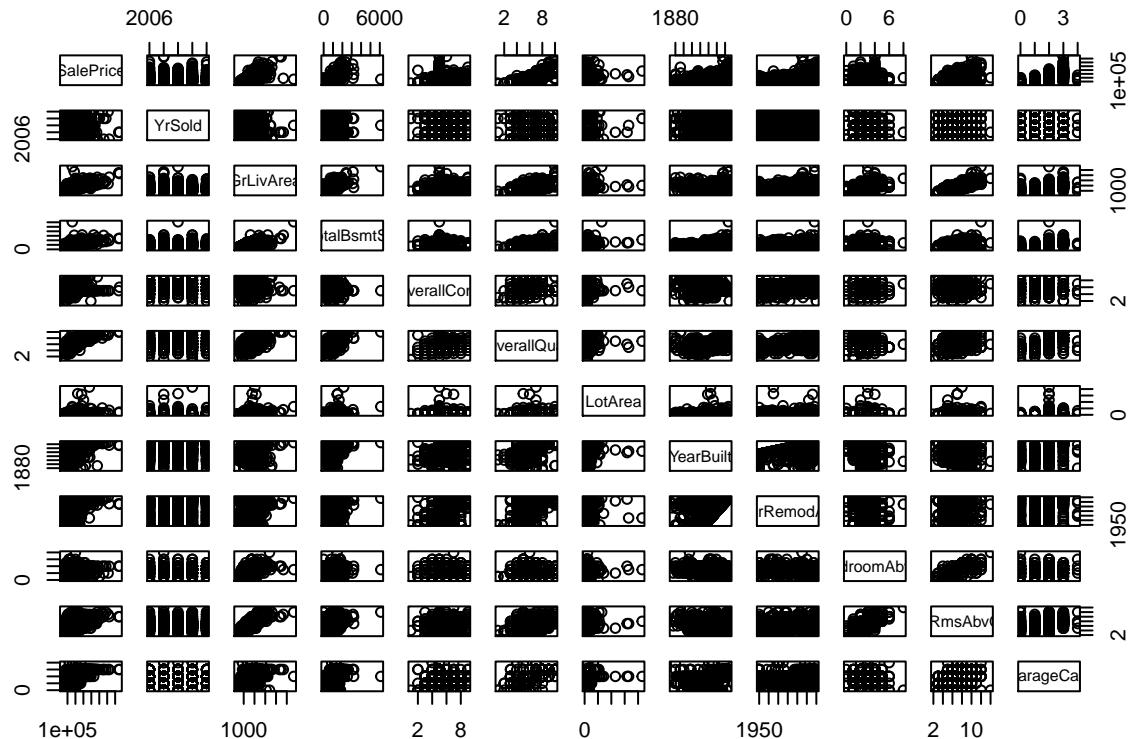
# Lab2

Michael St Jean

1/29/2020

## Exersize 1

Below is a visualization of our scatterplot matrix between 12 variables we believed to be associated with Sale Price. It is a little difficult to read in a pdf format, so if you are interested in getting a better view of each plot, run our rscript in github. The variables from left to right are sale price, year sold, total above ground living, total basement square footing, overall condition, overall quality, lot area, year built, year remodelled, bedrooms above ground, total rooms above ground, and garage size



In addition to these plots, we've included a matrix of correlation between the variables. The results can be seen below

```
##          SalePrice      YrSold   GrLivArea TotalBsmtSF OverallCond
## SalePrice 1.00000000 -0.02892259  0.70862448  0.61358055 -0.07785589
## YrSold    -0.02892259  1.00000000 -0.03652582 -0.01496865  0.04394975
## GrLivArea  0.70862448 -0.03652582  1.00000000  0.45486820 -0.07968587
## TotalBsmtSF 0.61358055 -0.01496865  0.45486820  1.00000000 -0.17109751
## OverallCond -0.07785589  0.04394975 -0.07968587 -0.17109751  1.00000000
## OverallQual  0.79098160 -0.02734671  0.59300743  0.53780850 -0.09193234
## LotArea     0.26384335 -0.01426141  0.26311617  0.26083313 -0.00563627
```

```

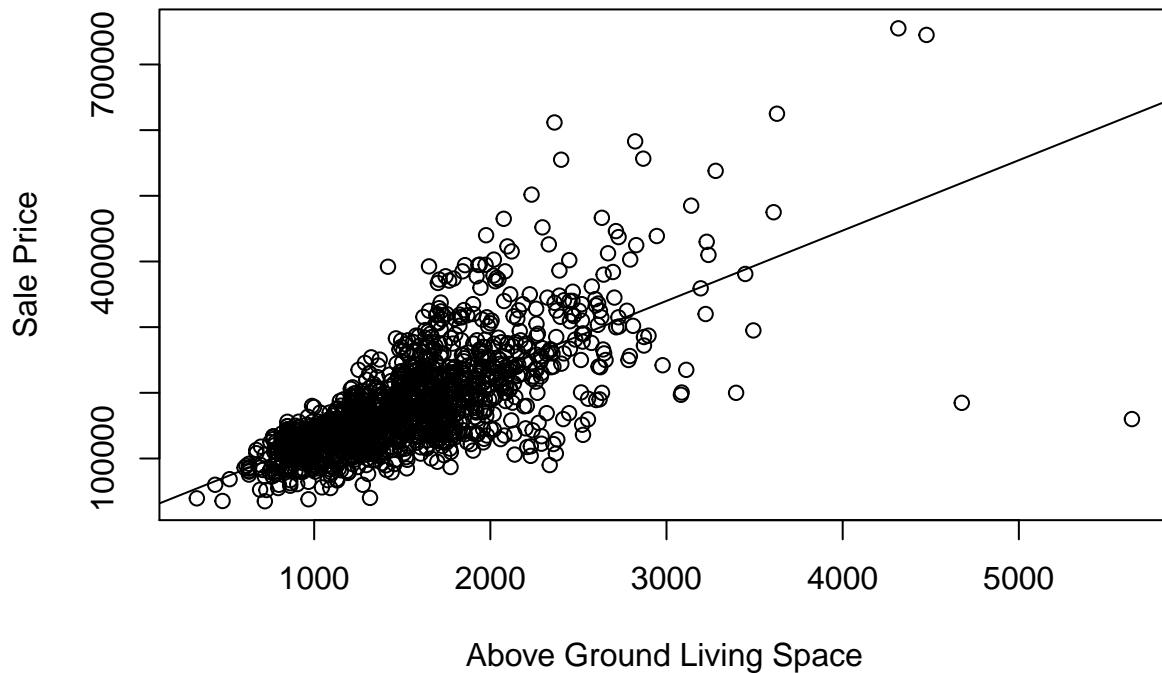
## YearBuilt      0.52289733 -0.01361768  0.19900971  0.39145200 -0.37598320
## YearRemodAdd  0.50710097  0.03574325  0.28738852  0.29106558  0.07374150
## BedroomAbvGr  0.16821315 -0.03601389  0.52126951  0.05044996  0.01298006
## TotRmsAbvGrd  0.53372316 -0.03451635  0.82548937  0.28557256 -0.05758317
## GarageCars    0.64040920 -0.03911690  0.46724742  0.43458483 -0.18575751
## OverallQual    OverallQual      LotArea   YearBuilt YearRemodAdd BedroomAbvGr
## SalePrice     0.79098160  0.26384335  0.52289733  0.50710097  0.16821315
## YrSold        -0.02734671 -0.01426141 -0.01361768  0.03574325 -0.03601389
## GrLivArea     0.59300743  0.26311617  0.19900971  0.28738852  0.52126951
## TotalBsmtSF   0.53780850  0.26083313  0.39145200  0.29106558  0.05044996
## OverallCond   -0.09193234 -0.00563627 -0.37598320  0.07374150  0.01298006
## OverallQual   1.00000000  0.10580574  0.57232277  0.55068392  0.10167636
## LotArea        0.10580574  1.00000000  0.01422765  0.01378843  0.11968991
## YearBuilt      0.57232277  0.01422765  1.00000000  0.59285498 -0.07065122
## YearRemodAdd  0.55068392  0.01378843  0.59285498  1.00000000 -0.04058093
## BedroomAbvGr  0.10167636  0.11968991 -0.07065122 -0.04058093  1.00000000
## TotRmsAbvGrd  0.42745234  0.19001478  0.09558913  0.19173982  0.67661994
## GarageCars    0.60067072  0.15487074  0.53785009  0.42062215  0.08610644
## TotRmsAbvGrd  TotRmsAbvGrd  GarageCars
## SalePrice     0.53372316  0.64040920
## YrSold        -0.03451635 -0.03911690
## GrLivArea     0.82548937  0.46724742
## TotalBsmtSF   0.28557256  0.43458483
## OverallCond   -0.05758317 -0.18575751
## OverallQual   0.42745234  0.60067072
## LotArea        0.19001478  0.15487074
## YearBuilt      0.09558913  0.53785009
## YearRemodAdd  0.19173982  0.42062215
## BedroomAbvGr  0.67661994  0.08610644
## TotRmsAbvGrd  1.00000000  0.36228857
## GarageCars    0.36228857  1.00000000

```

As we can see above the correlation factor relating each variable to one another is displayed, however, we are most concerned with the first column relating each variable to sale price. The variables with strong correlative relationships are above ground living area and overall quality. Variable with moderate correlative relationshios include total basement square footage, year built, year remodelled, total rooms above ground, garage size. The ones that didn't quite match our beliefs with weaker relationships were year sold, overall condition, lot area, and bedrooms above ground. Most of the correlations matched our earlier beliefs, all the variables that imply a bigger house are strongly correlated with an increase in sale price,newer and higher quality houses also sold for more, we expected condition to be positively correlated with sale price, we am unsure why this has a negative correlation, perhaps because condition is negatively correlated with basement size and YearBuilt. We expected Yrsold to be positively correlated with the saleprice but considering the timing of the dataset around the recession a negative correlation makes sense.

Now lets look at the relationship between sale price and above ground living space. Below us a scatterplot and simple linear regression model plotted alongside it.

## Sale Price vs Above Ground Living Space



```
## integer(0)
```

As we can see there are a few outliers to this notably the point

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.