

Linear Regression

Jieqi Tu (jt3098)

3/26/2019

Linear Regression

Load and tidy data

```
#read data
rawdata <- read.csv("kc_house_data.csv", header = TRUE)

#inspect the structure of data
str(rawdata)

## 'data.frame': 21613 obs. of 21 variables:
## $ id          : num  7.13e+09 6.41e+09 5.63e+09 2.49e+09 1.95e+09 ...
## $ date        : Factor w/ 372 levels "20140502T000000",...: 165 221 291 221 284 11 57 252 340 306 ...
## $ price       : num  221900 538000 180000 604000 510000 ...
## $ bedrooms    : int  3 3 2 4 3 4 3 3 3 3 ...
## $ bathrooms   : num  1 2.25 1 3 2 4.5 2.25 1.5 1 2.5 ...
## $ sqft_living : int  1180 2570 770 1960 1680 5420 1715 1060 1780 1890 ...
## $ sqft_lot    : int  5650 7242 10000 5000 8080 101930 6819 9711 7470 6560 ...
## $ floors      : num  1 2 1 1 1 1 2 1 1 2 ...
## $ waterfront  : int  0 0 0 0 0 0 0 0 0 ...
## $ view         : int  0 0 0 0 0 0 0 0 0 ...
## $ condition   : int  3 3 3 5 3 3 3 3 3 3 ...
## $ grade        : int  7 7 6 7 8 11 7 7 7 7 ...
## $ sqft_above   : int  1180 2170 770 1050 1680 3890 1715 1060 1050 1890 ...
## $ sqft_basement: int  0 400 0 910 0 1530 0 0 730 0 ...
## $ yr_built     : int  1955 1951 1933 1965 1987 2001 1995 1963 1960 2003 ...
## $ yr_renovated: int  0 1991 0 0 0 0 0 0 0 ...
## $ zipcode      : int  98178 98125 98028 98136 98074 98053 98003 98198 98146 98038 ...
## $ lat          : num  47.5 47.7 47.7 47.5 47.6 ...
## $ long         : num  -122 -122 -122 -122 -122 ...
## $ sqft_living15: int  1340 1690 2720 1360 1800 4760 2238 1650 1780 2390 ...
## $ sqft_lot15   : int  5650 7639 8062 5000 7503 101930 6819 9711 8113 7570 ...
```

`id, date, zipcode, lat, long` can be removed from the dataframe.

The cleaned dataset to be used in this project include one response variable `price` and additional 15 variables. `view`, `sqft_basement` and `yr_renovated` are continuous or integer variables in the original dataset. For the purpose of easy interpretation in later modelling, we convert these variables to be binary. Then these variables indicate whether the house has been viewed by potential buyers, whether the house has basement or not and whether the house has been renovated or not, respectively.

```
#clean the rawdata; create a tidied dataset for analysis and modelling.
```

```
housing =
  rawdata %>% select(-id, -date, -zipcode, -lat, -long) %>%
  mutate(view = ifelse(view == 0, 0, 1),
        basement = ifelse(sqft_basement == 0, 0, 1),
        renovated = ifelse(yr_renovated == 0, 0, 1)) %>%
  select(-sqft_basement, -yr_renovated)
```

Exploratory data analysis

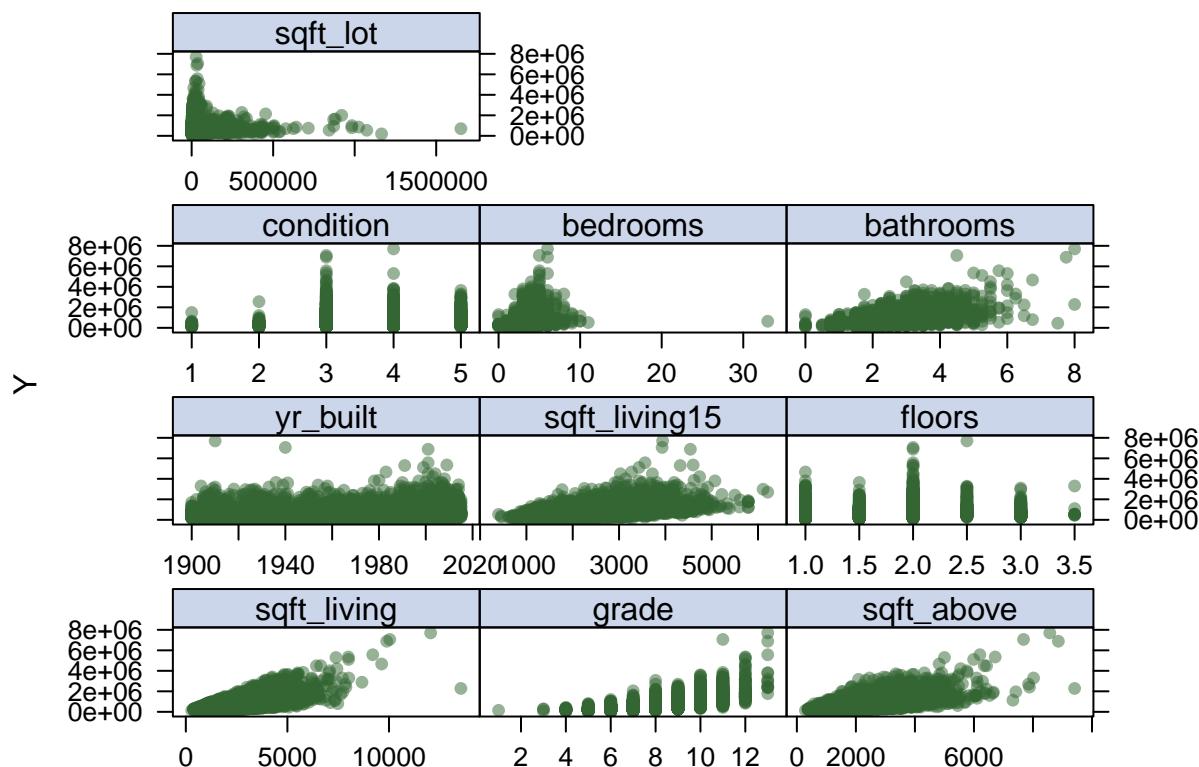
Continuous variables

```
# vector of response
y <- housing$price

# matrix of continuous predictors
housing_continious = housing %>% select(c(price, sqft_living, grade, sqft_above, yr_built, sqft_living15))
x_continuous <- model.matrix(price ~ ., housing_continious) [, -1]

# set up the theme of plotting
theme1 <- trellis.par.get()
theme1$plot.symbol$col <- rgb(.2, .4, .2, .5)
theme1$plot.symbol$pch <- 16
theme1$plot.line$col <- rgb(.8, .1, .1, 1)
theme1$plot.line$lwd <- 2
theme1$strip.background$col <- rgb(.0, .2, .6, .2)
trellis.par.set(theme1)

featurePlot(x_continuous, y, plot = "scatter", labels = c("", "Y"),
            type = c("p"), layout = c(3, 4))
```



Binary variables

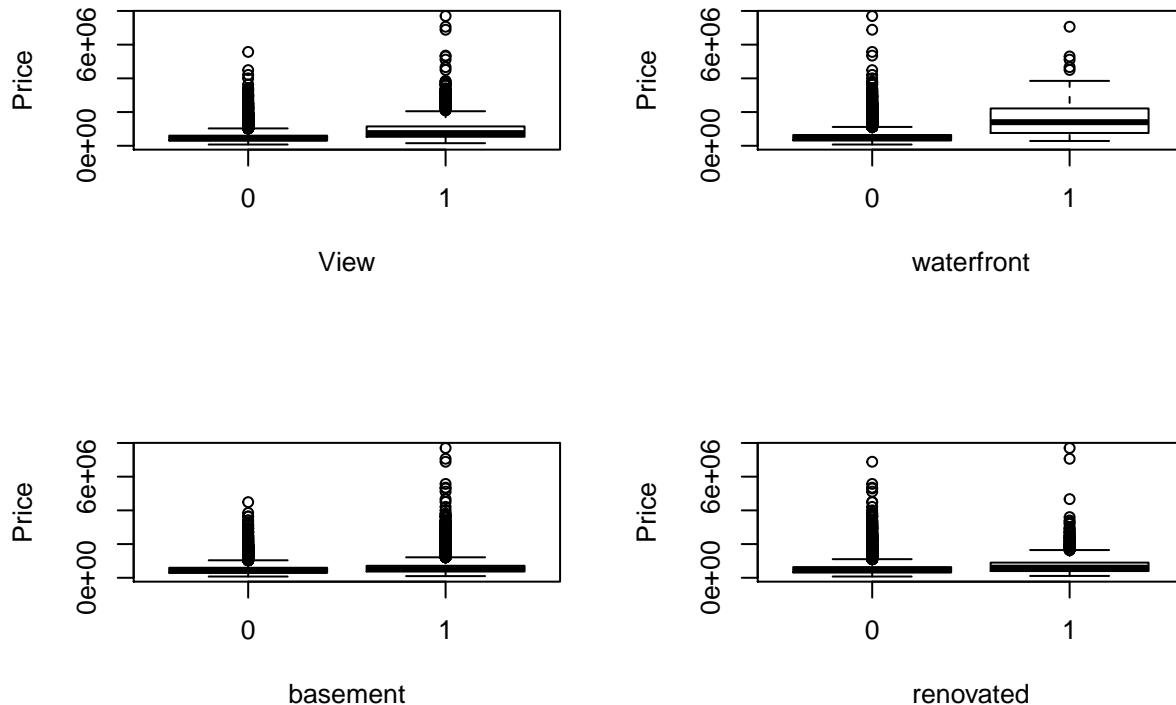
```
# data frame of binary predictors
housing_binary = housing %>%
```

```

select(c(price,view, waterfront, basement, renovated )) %>%
mutate(view = as.factor(view),
       waterfront = as.factor(waterfront),
       basement = as.factor(basement),
       renovated = as.factor(renovated))

# boxplots
par(mfrow = c(2, 2))
boxplot(price~view, data=housing_binary, ylab = "Price", xlab ="View")
boxplot(price~waterfront, data=housing_binary, ylab = "Price", xlab ="waterfront")
boxplot(price~basement, data=housing_binary, ylab = "Price", xlab ="basement")
boxplot(price~renovated, data=housing_binary, ylab = "Price", xlab ="renovated")

```



Model Building

Linear model

```

# build a least-square linear model
model_ls = lm(housing_binary$price~., data = housing_binary)
summary(model_ls)

```

```

##
## Call:
## lm(formula = housing_binary$price ~ ., data = housing_binary)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1400351 -190299  -70299  104701  6665645
##
## Coefficients:

```

```

##          Estimate Std. Error t value Pr(>|t|)
## (Intercept) 460299     2951  155.97 <2e-16 ***
## view1       347522     8044   43.20 <2e-16 ***
## waterfront1 765995    27180   28.18 <2e-16 ***
## basement1   86416     4725   18.29 <2e-16 ***
## renovated1  140118    11326   12.37 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 332500 on 21608 degrees of freedom
## Multiple R-squared:  0.1797, Adjusted R-squared:  0.1795
## F-statistic:  1183 on 4 and 21608 DF,  p-value: < 2.2e-16
# prepare the design matrix
x = model.matrix(price~., housing)[,-1]

# build a ridge model using glmnet function
model_ridge = glmnet(x, y, alpha = 0, lambda = exp(seq(-1, 15, length = 100)))

# check the coefficient matrix of the ridge model
coef(model_ridge)

## 16 x 100 sparse Matrix of class "dgCMatrix"
## [[ suppressing 100 column names 's0', 's1', 's2' ... ]]

##
## (Intercept) 2.849732e+05 2.884011e+05 3.013235e+05 3.258637e+05
## bedrooms    7.713687e+03 8.327204e+03 8.890343e+03 9.376145e+03
## bathrooms   1.773631e+04 1.955771e+04 2.142155e+04 2.330057e+04
## sqft_living 2.142085e+01 2.388851e+01 2.650037e+01 2.923933e+01
## sqft_lot    4.608417e-02 4.911242e-02 5.169891e-02 5.369467e-02
## floors      1.119644e+04 1.217683e+04 1.313790e+04 1.405986e+04
## waterfront  1.015216e+05 1.155365e+05 1.309794e+05 1.478708e+05
## view        3.652827e+04 4.108369e+04 4.597174e+04 5.116106e+04
## condition   3.523360e+03 4.215669e+03 5.019675e+03 5.942483e+03
## grade        1.621245e+04 1.814266e+04 2.020855e+04 2.240444e+04
## sqft_above   1.977818e+01 2.195160e+01 2.422543e+01 2.657993e+01
## yr_builtin -5.020667e+01 -7.226644e+01 -1.002132e+02 -1.349531e+02
## sqft_living15 2.332463e+01 2.590080e+01 2.859088e+01 3.136609e+01
## sqft_lot15   5.521419e-02 5.680746e-02 5.713271e-02 5.586475e-02
## basement    1.219363e+04 1.385244e+04 1.566579e+04 1.762924e+04
## renovated   2.076003e+04 2.358297e+04 2.666305e+04 2.998765e+04
##
## (Intercept) 3.641433e+05 4.181694e+05 4.897193e+05 5.802341e+05
## bedrooms    9.752111e+03 9.991728e+03 1.006237e+04 9.934924e+03
## bathrooms   2.516514e+04 2.698641e+04 2.873461e+04 3.038291e+04
## sqft_living 3.208807e+01 3.502215e+01 3.802013e+01 4.105955e+01
## sqft_lot    5.496959e-02 5.538773e-02 5.486345e-02 5.334704e-02
## floors      1.492692e+04 1.572077e+04 1.643151e+04 1.705315e+04
## waterfront  1.662018e+05 1.859259e+05 2.069578e+05 2.291702e+05
## view        5.660856e+04 6.225314e+04 6.802292e+04 7.383252e+04
## condition   6.986973e+03 8.153308e+03 9.434977e+03 1.081996e+04
## grade        2.472520e+04 2.716196e+04 2.970900e+04 3.236068e+04
## sqft_above   2.899379e+01 3.144363e+01 3.390762e+01 3.636502e+01
## yr_builtin -1.773687e+02 -2.282599e+02 -2.883054e+02 -3.580135e+02

```

```

## sqft_living15 3.419237e+01 3.703299e+01 3.984790e+01 4.259584e+01
## sqft_lot15 5.267323e-02 4.725989e-02 3.935809e-02 2.876457e-02
## basement 1.973334e+04 2.196280e+04 2.429714e+04 2.671076e+04
## renovated 3.353251e+04 3.726073e+04 4.112126e+04 4.504905e+04
##
## (Intercept) 6.907238e+05 8.217011e+05 9.734733e+05 1.145077e+06
## bedrooms 9.583371e+03 8.986210e+03 8.125862e+03 7.004421e+03
## bathrooms 3.190794e+04 3.329077e+04 3.453980e+04 3.562438e+04
## sqft_living 4.411893e+01 4.717899e+01 5.024243e+01 5.326521e+01
## sqft_lot 5.083962e-02 4.739981e-02 4.318002e-02 3.826529e-02
## floors 1.758537e+04 1.803360e+04 1.842223e+04 1.874091e+04
## waterfront 2.523931e+05 2.764172e+05 3.010097e+05 3.258895e+05
## view 7.958637e+04 8.518240e+04 9.052473e+04 9.549852e+04
## condition 1.229015e+04 1.382156e+04 1.538106e+04 1.694322e+04
## grade 3.511284e+04 3.796318e+04 4.091145e+04 4.395167e+04
## sqft_above 3.879756e+01 4.118989e+01 4.351748e+01 4.578207e+01
## yr_built -4.376844e+02 -5.273849e+02 -6.271275e+02 -7.362371e+02
## sqft_living15 4.523521e+01 4.772488e+01 5.001596e+01 5.208419e+01
## sqft_lot15 1.535367e-02 -9.121298e-04 -2.001338e-02 -4.172024e-02
## basement 2.917399e+04 3.165455e+04 3.411407e+04 3.652428e+04
## renovated 4.896614e+04 5.278390e+04 5.640084e+04 5.972638e+04
##
## (Intercept) 1.335583e+06 1.544031e+06 1.767493e+06 2.003988e+06
## bedrooms 5.618432e+03 3.978548e+03 2.110412e+03 3.640898e+01
## bathrooms 3.655191e+04 3.736561e+04 3.802507e+04 3.856268e+04
## sqft_living 5.624942e+01 5.921342e+01 6.210833e+01 6.495660e+01
## sqft_lot 3.289670e-02 2.732374e-02 2.172073e-02 1.636687e-02
## floors 1.901476e+04 1.927502e+04 1.950473e+04 1.973040e+04
## waterfront 3.507805e+05 3.754116e+05 3.994950e+05 4.227897e+05
## view 1.000184e+05 1.040162e+05 1.074195e+05 1.102033e+05
## condition 1.847061e+04 1.992392e+04 2.128020e+04 2.250628e+04
## grade 4.708708e+04 5.031139e+04 5.362435e+04 5.702395e+04
## sqft_above 4.796971e+01 5.004587e+01 5.203472e+01 5.390944e+01
## yr_built -8.541488e+02 -9.803052e+02 -1.113126e+03 -1.251596e+03
## sqft_living15 5.388839e+01 5.538546e+01 5.656763e+01 5.740064e+01
## sqft_lot15 -6.582195e-02 -9.201871e-02 -1.198555e-01 -1.489595e-01
## basement 3.885233e+04 4.105783e+04 4.313169e+04 4.504613e+04
## renovated 6.266280e+04 6.511363e+04 6.702491e+04 6.833802e+04
##
## (Intercept) 2.251613e+06 2.506069e+06 2.765564e+06 3.025765e+06
## bedrooms -2.198254e+03 -4.564082e+03 -7.002969e+03 -9.487513e+03
## bathrooms 3.906286e+04 3.944029e+04 3.982823e+04 4.012838e+04
## sqft_living 6.777594e+01 7.053597e+01 7.326449e+01 7.596131e+01
## sqft_lot 1.138529e-02 7.026912e-03 3.251740e-03 3.074455e-04
## floors 1.996545e+04 2.019478e+04 2.042701e+04 2.065676e+04
## waterfront 4.450923e+05 4.662062e+05 4.860154e+05 5.044122e+05
## view 1.123619e+05 1.139028e+05 1.148680e+05 1.153107e+05
## condition 2.357699e+04 2.448282e+04 2.521113e+04 2.576325e+04
## grade 6.047633e+04 6.400423e+04 6.754527e+04 7.113048e+04
## sqft_above 5.561477e+01 5.719589e+01 5.856745e+01 5.976361e+01
## yr_built -1.394699e+03 -1.540290e+03 -1.687399e+03 -1.833927e+03
## sqft_living15 5.787021e+01 5.797285e+01 5.772006e+01 5.710753e+01
## sqft_lot15 -1.787749e-01 -2.089572e-01 -2.388482e-01 -2.682782e-01
## basement 4.675894e+04 4.828955e+04 4.958366e+04 5.066542e+04

```

```

## renovated      6.901166e+04  6.906703e+04  6.850439e+04  6.738810e+04
##
## (Intercept) 3.284516e+06  3.538312e+06  3.784262e+06  4.020305e+06
## bedrooms     -1.195378e+04 -1.437457e+04 -1.672421e+04 -1.895477e+04
## bathrooms    4.047866e+04  4.082548e+04  4.113067e+04  4.150065e+04
## sqft_living   7.862245e+01  8.126448e+01  8.392848e+01  8.656809e+01
## sqft_lot      -2.058164e-03 -3.747686e-03 -4.699560e-03 -5.301570e-03
## floors        2.087011e+04  2.106706e+04  2.125598e+04  2.140969e+04
## waterfront   5.213609e+05  5.368394e+05  5.508619e+05  5.634821e+05
## view          1.152969e+05  1.149023e+05  1.142048e+05  1.132738e+05
## condition    2.614578e+04  2.637003e+04  2.644997e+04  2.641004e+04
## grade         7.467300e+04  7.817906e+04  8.164891e+04  8.499040e+04
## sqft_above    6.071137e+01  6.141384e+01  6.184643e+01  6.198739e+01
## yr_builtin   -1.978689e+03 -2.119985e+03 -2.256429e+03 -2.386874e+03
## sqft_living15 5.619379e+01  5.499682e+01  5.353499e+01  5.191128e+01
## sqft_lot15   -2.965071e-01 -3.233848e-01 -3.488255e-01 -3.722227e-01
## basement     5.148758e+04  5.206014e+04  5.238015e+04  5.243244e+04
## renovated     6.576873e+04  6.373259e+04  6.136645e+04  5.876099e+04
##
## (Intercept) 4.244367e+06  4.454629e+06  4.651047e+06  4.832753e+06
## bedrooms     -2.106688e+04 -2.302303e+04 -2.485183e+04 -2.652699e+04
## bathrooms    4.184139e+04  4.223355e+04  4.258461e+04  4.298755e+04
## sqft_living   8.928045e+01  9.197038e+01  9.485703e+01  9.781752e+01
## sqft_lot      -5.352430e-03 -5.264602e-03 -4.710219e-03 -4.129871e-03
## floors        2.156158e+04  2.167510e+04  2.182370e+04  2.196681e+04
## waterfront   5.747540e+05  5.847674e+05  5.935930e+05  6.013363e+05
## view          1.121797e+05  1.109778e+05  1.097171e+05  1.084328e+05
## condition    2.626344e+04  2.604123e+04  2.574335e+04  2.539866e+04
## grade         8.823937e+04  9.130880e+04  9.425710e+04  9.699569e+04
## sqft_above    6.179626e+01  6.131407e+01  6.039605e+01  5.911179e+01
## yr_builtin   -2.510418e+03 -2.626040e+03 -2.733899e+03 -2.833443e+03
## sqft_living15 5.011939e+01  4.827743e+01  4.634061e+01  4.443314e+01
## sqft_lot15   -3.938613e-01 -4.132608e-01 -4.309804e-01 -4.465831e-01
## basement     5.221296e+04  5.174602e+04  5.096797e+04  4.991770e+04
## renovated     5.599646e+04  5.316967e+04  5.031341e+04  4.750442e+04
##
## (Intercept) 4.999119e+06  5.148281e+06  5.283776e+06  5.403210e+06
## bedrooms     -2.804418e+04 -2.939492e+04 -3.061123e+04 -3.169041e+04
## bathrooms    4.343307e+04  4.381460e+04  4.426784e+04  4.458865e+04
## sqft_living   1.008089e+02  1.037446e+02  1.067371e+02  1.097469e+02
## sqft_lot      -3.625160e-03 -3.144287e-03 -2.819292e-03 -2.376293e-03
## floors        2.209835e+04  2.220230e+04  2.232141e+04  2.244726e+04
## waterfront   6.080973e+05  6.139798e+05  6.190657e+05  6.234474e+05
## view          1.071561e+05  1.059180e+05  1.047250e+05  1.035975e+05
## condition    2.502749e+04  2.465857e+04  2.427851e+04  2.391487e+04
## grade         9.950740e+04  1.018168e+05  1.038831e+05  1.057891e+05
## sqft_above    5.751578e+01  5.575448e+01  5.368530e+01  5.145873e+01
## yr_builtin   -2.924382e+03 -3.005889e+03 -3.079706e+03 -3.144838e+03
## sqft_living15 4.260802e+01  4.088345e+01  3.929547e+01  3.779699e+01
## sqft_lot15   -4.601170e-01 -4.718998e-01 -4.818610e-01 -4.906628e-01
## basement     4.863854e+04  4.724054e+04  4.563275e+04  4.393558e+04
## renovated     4.480168e+04  4.228428e+04  3.990919e+04  3.774959e+04
##
## (Intercept) 5.509032e+06  5.601448e+06  5.682735e+06  5.754454e+06

```

```

## bedrooms      -3.264823e+04 -3.349983e+04 -3.425417e+04 -3.492302e+04
## bathrooms     4.487851e+04  4.504360e+04  4.518862e+04  4.535935e+04
## sqft_living   1.127647e+02  1.158722e+02  1.190033e+02  1.220859e+02
## sqft_lot       -1.983485e-03 -1.438932e-03 -9.402479e-04 -5.782680e-04
## floors        2.258653e+04  2.276043e+04  2.296128e+04  2.317565e+04
## waterfront    6.272067e+05  6.304130e+05  6.331394e+05  6.354511e+05
## view          1.025373e+05  1.015456e+05  1.006203e+05  9.976180e+04
## condition     2.356460e+04  2.323233e+04  2.291612e+04  2.261653e+04
## grade          1.074994e+05  1.090721e+05  1.104684e+05  1.116811e+05
## sqft_above     4.906103e+01  4.647327e+01  4.372993e+01  4.090860e+01
## yr_builtin    -3.202490e+03  -3.252908e+03  -3.297198e+03  -3.336159e+03
## sqft_living15 3.642903e+01  3.514348e+01  3.398909e+01  3.298021e+01
## sqft_lot15    -4.981995e-01  -5.049376e-01  -5.106942e-01  -5.154706e-01
## basement      4.213674e+04  4.023912e+04  3.825736e+04  3.623625e+04
## renovated     3.578163e+04  3.401189e+04  3.241590e+04  3.097796e+04
##
## (Intercept)  5.816513e+06  5.870710e+06  5.917377e+06  5.957845e+06
## bedrooms     -3.551710e+04  -3.604086e+04  -3.650588e+04  -3.691436e+04
## bathrooms    4.544120e+04  4.553605e+04  4.556448e+04  4.560191e+04
## sqft_living   1.252322e+02  1.282563e+02  1.312877e+02  1.341411e+02
## sqft_lot      -1.039540e-04  2.545339e-04  6.949177e-04  1.039602e-03
## floors        2.342819e+04  2.368128e+04  2.396212e+04  2.423291e+04
## waterfront   6.373966e+05  6.390370e+05  6.404066e+05  6.415562e+05
## view          9.896420e+04  9.823214e+04  9.755531e+04  9.694028e+04
## condition     2.233590e+04  2.207621e+04  2.183497e+04  2.161527e+04
## grade          1.127799e+05  1.137279e+05  1.145807e+05  1.153127e+05
## sqft_above     3.795769e+01  3.504505e+01  3.208045e+01  2.924242e+01
## yr_builtin    -3.369903e+03  -3.399292e+03  -3.424612e+03  -3.446513e+03
## sqft_living15 3.205338e+01  3.125303e+01  3.052568e+01  2.990239e+01
## sqft_lot15    -5.197449e-01  -5.232906e-01  -5.264688e-01  -5.291103e-01
## basement      3.415824e+04  3.211740e+04  3.006699e+04  2.811015e+04
## renovated     2.969986e+04  2.856428e+04  2.756087e+04  2.667774e+04
##
## (Intercept)  5.992691e+06  6.022626e+06  6.048394e+06  6.070383e+06
## bedrooms     -3.727394e+04  -3.759014e+04  -3.786424e+04  -3.810682e+04
## bathrooms    4.562085e+04  4.562252e+04  4.563817e+04  4.562054e+04
## sqft_living   1.368560e+02  1.394216e+02  1.417263e+02  1.439330e+02
## sqft_lot      1.363778e-03  1.674051e-03  1.893994e-03  2.159918e-03
## floors        2.450160e+04  2.476537e+04  2.500171e+04  2.524076e+04
## waterfront   6.425159e+05  6.433153e+05  6.439873e+05  6.445414e+05
## view          9.638119e+04  9.587470e+04  9.542496e+04  9.501669e+04
## condition     2.141540e+04  2.123434e+04  2.107424e+04  2.092861e+04
## grade          1.159516e+05  1.165095e+05  1.169802e+05  1.174007e+05
## sqft_above     2.651118e+01  2.390680e+01  2.154411e+01  1.927442e+01
## yr_builtin    -3.465349e+03  -3.481514e+03  -3.495390e+03  -3.507240e+03
## sqft_living15 2.935735e+01  2.888067e+01  2.848171e+01  2.812203e+01
## sqft_lot15    -5.313821e-01  -5.333457e-01  -5.349467e-01  -5.364031e-01
## basement      2.623757e+04  2.446128e+04  2.284952e+04  2.131297e+04
## renovated     2.590337e+04  2.522608e+04  2.463865e+04  2.412530e+04
##
## (Intercept)  6.089244e+06  6.105316e+06  6.119060e+06  6.130773e+06
## bedrooms     -3.831557e+04  -3.850078e+04  -3.865869e+04  -3.879925e+04
## bathrooms    4.561614e+04  4.558584e+04  4.556965e+04  4.553441e+04
## sqft_living   1.458802e+02  1.477307e+02  1.493317e+02  1.508437e+02

```

```

## sqft_lot      2.351826e-03  2.585138e-03  2.752174e-03  2.953136e-03
## floors       2.545053e+04  2.565989e+04  2.583918e+04  2.601607e+04
## waterfront   6.450081e+05  6.453905e+05  6.457142e+05  6.459780e+05
## view          9.465809e+04  9.433374e+04  9.405227e+04  9.379841e+04
## condition    2.080143e+04  2.068592e+04  2.058643e+04  2.049609e+04
## grade         1.177540e+05  1.180711e+05  1.183355e+05  1.185738e+05
## sqft_above    1.725667e+01  1.533733e+01  1.366673e+01  1.209005e+01
## yr_builtin   -3.517377e+03 -3.526025e+03  -3.533401e+03 -3.539695e+03
## sqft_living15 2.782297e+01  2.755123e+01  2.732765e+01  2.712286e+01
## sqft_lot15   -5.375882e-01 -5.386825e-01  -5.395658e-01 -5.403911e-01
## basement     1.994604e+04  1.865479e+04  1.752934e+04  1.647382e+04
## renovated    2.368294e+04  2.329647e+04  2.296575e+04  2.267652e+04
##
## (Intercept)  6.140730e+06  6.149278e+06  6.156483e+06  6.162593e+06
## bedrooms     -3.891255e+04 -3.901711e+04  -3.910404e+04 -3.917736e+04
## bathrooms    4.553363e+04  4.550936e+04  4.549164e+04  4.547993e+04
## sqft_living   1.520201e+02  1.531784e+02  1.541542e+02  1.549775e+02
## sqft_lot      3.032033e-03  3.170901e-03  3.272914e-03  3.346095e-03
## floors        2.614519e+04  2.628085e+04  2.639427e+04  2.648874e+04
## waterfront   6.462115e+05  6.463973e+05  6.465566e+05  6.466926e+05
## view          9.358964e+04  9.339767e+04  9.323449e+04  9.309572e+04
## condition    2.042372e+04  2.035563e+04  2.029869e+04  2.025075e+04
## grade         1.187571e+05  1.189297e+05  1.190724e+05  1.191909e+05
## sqft_above    1.085050e+01  9.636727e+00  8.609880e+00  7.741232e+00
## yr_builtin   -3.545019e+03 -3.549600e+03  -3.553457e+03 -3.556722e+03
## sqft_living15 2.697266e+01  2.682484e+01  2.670540e+01  2.660748e+01
## sqft_lot15   -5.409675e-01 -5.415539e-01  -5.420272e-01 -5.424119e-01
## basement     1.563679e+04  1.482440e+04  1.413660e+04  1.355387e+04
## renovated    2.243729e+04  2.222386e+04  2.204471e+04  2.189333e+04
##
## (Intercept)  6.167901e+06  6.172276e+06  6.175938e+06  6.179248e+06
## bedrooms     -3.924531e+04 -3.929977e+04  -3.934403e+04 -3.938651e+04
## bathrooms    4.545722e+04  4.544087e+04  4.543270e+04  4.542087e+04
## sqft_living   1.557791e+02  1.564233e+02  1.569352e+02  1.574429e+02
## sqft_lot      3.452825e-03  3.523312e-03  3.560905e-03  3.621275e-03
## floors        2.658567e+04  2.666195e+04  2.672009e+04  2.678123e+04
## waterfront   6.468002e+05  6.468948e+05  6.469775e+05  6.470428e+05
## view          9.296812e+04  9.286319e+04  9.277725e+04  9.269644e+04
## condition    2.020530e+04  2.016922e+04  2.014048e+04  2.011168e+04
## grade         1.193026e+05  1.193917e+05  1.194624e+05  1.195306e+05
## sqft_above    6.898965e+00  6.219661e+00  5.677976e+00  5.142082e+00
## yr_builtin   -3.559565e+03 -3.561906e+03  -3.563860e+03 -3.565628e+03
## sqft_living15 2.651105e+01  2.643695e+01  2.638002e+01  2.632152e+01
## sqft_lot15   -5.427968e-01 -5.430969e-01  -5.433249e-01 -5.435529e-01
## basement     1.299276e+04  1.253927e+04  1.217587e+04  1.181855e+04
## renovated    2.175691e+04  2.164609e+04  2.155510e+04  2.146974e+04
##
## (Intercept)  6.181823e+06  6.184222e+06  6.186309e+06  6.187786e+06
## bedrooms     -3.941840e+04 -3.944930e+04  -3.947803e+04 -3.949715e+04
## bathrooms    4.541239e+04  4.540353e+04  4.538985e+04  4.538085e+04
## sqft_living   1.578199e+02  1.581942e+02  1.585585e+02  1.587972e+02
## sqft_lot      3.649836e-03  3.693887e-03  3.752799e-03  3.774716e-03
## floors        2.682397e+04  2.686929e+04  2.691556e+04  2.694279e+04
## waterfront   6.471007e+05  6.471478e+05  6.471841e+05  6.472177e+05

```

```

## view          9.263446e+04  9.257501e+04  9.252023e+04  9.248322e+04
## condition    2.009136e+04  2.007023e+04  2.005016e+04  2.003840e+04
## grade         1.195816e+05  1.196310e+05  1.196781e+05  1.197096e+05
## sqft_above    4.743800e+00  4.348056e+00  3.964668e+00  3.713701e+00
## yr_built     -3.567003e+03 -3.568284e+03 -3.569401e+03 -3.570193e+03
## sqft_living15 2.628040e+01  2.623823e+01  2.619633e+01  2.617067e+01
## sqft_lot15   -5.437224e-01 -5.438868e-01 -5.440540e-01 -5.441663e-01
## basement      1.155161e+04  1.128792e+04  1.103419e+04  1.086646e+04
## renovated     2.140516e+04  2.134306e+04  2.128685e+04  2.124890e+04
##
## (Intercept)  6.189238e+06  6.190546e+06  6.191686e+06  6.192662e+06
## bedrooms     -3.951609e+04 -3.953410e+04 -3.955085e+04 -3.956620e+04
## bathrooms     4.537371e+04  4.536513e+04  4.535407e+04  4.534065e+04
## sqft_living   1.590336e+02  1.592655e+02  1.594901e+02  1.597054e+02
## sqft_lot       3.804203e-03  3.842113e-03  3.887432e-03  3.938377e-03
## floors        2.697179e+04  2.700159e+04  2.703159e+04  2.706134e+04
## waterfront    6.472464e+05  6.472686e+05  6.472853e+05  6.472976e+05
## view          9.244596e+04  9.241107e+04  9.237888e+04  9.234940e+04
## condition     2.002537e+04  2.001252e+04  2.000035e+04  1.998903e+04
## grade         1.197402e+05  1.197697e+05  1.197978e+05  1.198243e+05
## sqft_above    3.463701e+00  3.219085e+00  2.983036e+00  2.757785e+00
## yr_built     -3.570968e+03 -3.571668e+03 -3.572280e+03 -3.572807e+03
## sqft_living15 2.614485e+01  2.611873e+01  2.609291e+01  2.606790e+01
## sqft_lot15   -5.442702e-01 -5.443737e-01 -5.444783e-01 -5.445830e-01
## basement      1.070032e+04  1.053863e+04  1.038352e+04  1.023638e+04
## renovated     2.121102e+04  2.117562e+04  2.114351e+04  2.111480e+04
##
## (Intercept)  6.193491e+06  6.194194e+06  6.194788e+06  6.195290e+06
## bedrooms     -3.958010e+04 -3.959261e+04 -3.960383e+04 -3.961387e+04
## bathrooms     4.532542e+04  4.530904e+04  4.529209e+04  4.527504e+04
## sqft_living   1.599096e+02  1.601017e+02  1.602810e+02  1.604475e+02
## sqft_lot       3.992956e-03  4.049347e-03  4.106046e-03  4.161893e-03
## floors        2.709041e+04  2.711845e+04  2.714522e+04  2.717052e+04
## waterfront    6.473064e+05  6.473125e+05  6.473165e+05  6.473188e+05
## view          9.232253e+04  9.229812e+04  9.227600e+04  9.225601e+04
## condition     1.997862e+04  1.996910e+04  1.996042e+04  1.995255e+04
## grade         1.198492e+05  1.198723e+05  1.198937e+05  1.199134e+05
## sqft_above    2.544815e+00  2.345009e+00  2.158781e+00  1.986185e+00
## yr_built     -3.573256e+03 -3.573639e+03 -3.573966e+03 -3.574243e+03
## sqft_living15 2.604405e+01  2.602157e+01  2.600056e+01  2.598105e+01
## sqft_lot15   -5.446861e-01 -5.447862e-01 -5.448820e-01 -5.449727e-01
## basement      1.009802e+04  9.968832e+03  9.848939e+03  9.738232e+03
## renovated     2.108932e+04  2.106677e+04  2.104685e+04  2.102926e+04
##
## (Intercept)  6.195715e+06  6.196075e+06  6.196379e+06  6.196636e+06
## bedrooms     -3.962283e+04 -3.963082e+04 -3.963794e+04 -3.964428e+04
## bathrooms     4.525826e+04  4.524201e+04  4.522645e+04  4.521172e+04
## sqft_living   1.606013e+02  1.607426e+02  1.608721e+02  1.609902e+02
## sqft_lot       4.216034e-03  4.267869e-03  4.317000e-03  4.363185e-03
## floors        2.719425e+04  2.721637e+04  2.723686e+04  2.725576e+04
## waterfront    6.473198e+05  6.473199e+05  6.473193e+05  6.473182e+05
## view          9.223797e+04  9.222173e+04  9.220713e+04  9.219402e+04
## condition     1.994541e+04  1.993896e+04  1.993314e+04  1.992790e+04
## grade         1.199315e+05  1.199480e+05  1.199631e+05  1.199768e+05

```

```

## sqft_above      1.827003e+00  1.680817e+00  1.547072e+00  1.425115e+00
## yr_built      -3.574479e+03 -3.574680e+03 -3.574851e+03 -3.574997e+03
## sqft_living15   2.596304e+01  2.594649e+01  2.593133e+01  2.591750e+01
## sqft_lot15     -5.450580e-01 -5.451375e-01 -5.452113e-01 -5.452793e-01
## basement       9.636457e+03  9.543256e+03  9.458199e+03  9.380813e+03
## renovated      2.101375e+04  2.100006e+04  2.098799e+04  2.097734e+04
##
## (Intercept)    6.196853e+06  6.197037e+06  6.197192e+06  6.197324e+06
## bedrooms      -3.964992e+04 -3.965492e+04 -3.965937e+04 -3.966332e+04
## bathrooms      4.519788e+04  4.518496e+04  4.517298e+04  4.516192e+04
## sqft_living    1.610977e+02  1.611952e+02  1.612835e+02  1.613632e+02
## sqft_lot       4.406301e-03  4.446318e-03  4.483270e-03  4.517242e-03
## floors         2.727311e+04  2.728898e+04  2.730345e+04  2.731660e+04
## waterfront     6.473167e+05  6.473149e+05  6.473131e+05  6.473111e+05
## view           9.218228e+04  9.217178e+04  9.216240e+04  9.215402e+04
## condition      1.992319e+04  1.991897e+04  1.991518e+04  1.991179e+04
## grade          1.199892e+05  1.200004e+05  1.200105e+05  1.200196e+05
## sqft_above      1.314241e+00  1.213713e+00  1.122787e+00  1.040728e+00
## yr_built      -3.575122e+03 -3.575227e+03 -3.575318e+03 -3.575395e+03
## sqft_living15   2.590491e+01  2.589349e+01  2.588315e+01  2.587381e+01
## sqft_lot15     -5.453419e-01 -5.453991e-01 -5.454513e-01 -5.454988e-01
## basement       9.310599e+03  9.247051e+03  9.189667e+03  9.137956e+03
## renovated      2.096796e+04  2.095970e+04  2.095242e+04  2.094601e+04
##
## (Intercept)    6.197435e+06  6.197529e+06  6.197608e+06  6.197675e+06
## bedrooms      -3.966682e+04 -3.966991e+04 -3.967266e+04 -3.967508e+04
## bathrooms      4.515175e+04  4.514245e+04  4.513396e+04  4.512624e+04
## sqft_living    1.614350e+02  1.614997e+02  1.615577e+02  1.616097e+02
## sqft_lot       4.548354e-03  4.576748e-03  4.602582e-03  4.626021e-03
## floors         2.732852e+04  2.733930e+04  2.734903e+04  2.735780e+04
## waterfront     6.473091e+05  6.473072e+05  6.473053e+05  6.473035e+05
## view           9.214655e+04  9.213990e+04  9.213399e+04  9.212873e+04
## condition      1.990876e+04  1.990606e+04  1.990365e+04  1.990151e+04
## grade          1.200278e+05  1.200351e+05  1.200417e+05  1.200476e+05
## sqft_above      9.668212e-01  9.003797e-01  8.407517e-01  7.873225e-01
## yr_built      -3.575460e+03 -3.575516e+03 -3.575563e+03 -3.575604e+03
## sqft_living15   2.586540e+01  2.585783e+01  2.585103e+01  2.584493e+01
## sqft_lot15     -5.455419e-01 -5.455809e-01 -5.456161e-01 -5.456478e-01
## basement       9.091446e+03  9.049686e+03  9.012252e+03  8.978745e+03
## renovated      2.094037e+04  2.093541e+04  2.093105e+04  2.092721e+04
##
## (Intercept)    6.197731e+06  6.197779e+06  6.197819e+06  6.197853e+06
## bedrooms      -3.967723e+04 -3.967912e+04 -3.968080e+04 -3.968227e+04
## bathrooms      4.511923e+04  4.511289e+04  4.510717e+04  4.510201e+04
## sqft_living    1.616562e+02  1.616978e+02  1.617350e+02  1.617681e+02
## sqft_lot       4.647235e-03  4.666390e-03  4.683652e-03  4.699177e-03
## floors         2.736567e+04  2.737274e+04  2.737907e+04  2.738473e+04
## waterfront     6.473018e+05  6.473002e+05  6.472987e+05  6.472974e+05
## view           9.212406e+04  9.211992e+04  9.211625e+04  9.211300e+04
## condition      1.989960e+04  1.989791e+04  1.989640e+04  1.989507e+04
## grade          1.200529e+05  1.200576e+05  1.200617e+05  1.200655e+05
## sqft_above      7.395176e-01  6.968029e-01  6.586844e-01  6.247076e-01
## yr_built      -3.575639e+03 -3.575668e+03 -3.575693e+03 -3.575714e+03
## sqft_living15   2.583947e+01  2.583458e+01  2.583022e+01  2.582634e+01

```

```

## sqft_lot15 -5.456763e-01 -5.457019e-01 -5.457249e-01 -5.457454e-01
## basement 8.948795e+03 8.922058e+03 8.898219e+03 8.876987e+03
## renovated 2.092385e+04 2.092089e+04 2.091829e+04 2.091600e+04
##
## (Intercept) 6.197882e+06 6.197906e+06 6.197926e+06 6.197943e+06
## bedrooms -3.968358e+04 -3.968473e+04 -3.968574e+04 -3.968663e+04
## bathrooms 4.509738e+04 4.509322e+04 4.508949e+04 4.508615e+04
## sqft_living 1.617976e+02 1.618238e+02 1.618471e+02 1.618678e+02
## sqft_lot 4.713115e-03 4.725610e-03 4.736793e-03 4.746788e-03
## floors 2.738979e+04 2.739431e+04 2.739833e+04 2.740191e+04
## waterfront 6.472961e+05 6.472949e+05 6.472938e+05 6.472928e+05
## view 9.211012e+04 9.210758e+04 9.210533e+04 9.210334e+04
## condition 1.989388e+04 1.989283e+04 1.989191e+04 1.989108e+04
## grade 1.200688e+05 1.200717e+05 1.200744e+05 1.200767e+05
## sqft_above 5.944555e-01 5.675476e-01 5.436373e-01 5.224099e-01
## yr_built -3.575733e+03 -3.575748e+03 -3.575761e+03 -3.575773e+03
## sqft_living15 2.582287e+01 2.581979e+01 2.581704e+01 2.581461e+01
## sqft_lot15 -5.457638e-01 -5.457802e-01 -5.457948e-01 -5.458079e-01
## basement 8.858096e+03 8.841306e+03 8.826396e+03 8.813167e+03
## renovated 2.091400e+04 2.091224e+04 2.091070e+04 2.090935e+04
##
## (Intercept) 6.197958e+06 6.197970e+06 6.197980e+06 6.197989e+06
## bedrooms -3.968742e+04 -3.968811e+04 -3.968871e+04 -3.968925e+04
## bathrooms 4.508316e+04 4.508050e+04 4.507813e+04 4.507601e+04
## sqft_living 1.618862e+02 1.619025e+02 1.619169e+02 1.619297e+02
## sqft_lot 4.755711e-03 4.763666e-03 4.770751e-03 4.777054e-03
## floors 2.740510e+04 2.740793e+04 2.741044e+04 2.741267e+04
## waterfront 6.472920e+05 6.472911e+05 6.472904e+05 6.472898e+05
## view 9.210158e+04 9.210003e+04 9.209867e+04 9.209746e+04
## condition 1.989036e+04 1.988972e+04 1.988915e+04 1.988865e+04
## grade 1.200787e+05 1.200806e+05 1.200822e+05 1.200836e+05
## sqft_above 5.035803e-01 4.868911e-01 4.721102e-01 4.590288e-01
## yr_built -3.575782e+03 -3.575790e+03 -3.575797e+03 -3.575803e+03
## sqft_living15 2.581245e+01 2.581053e+01 2.580883e+01 2.580733e+01
## sqft_lot15 -5.458194e-01 -5.458297e-01 -5.458389e-01 -5.458470e-01
## basement 8.801439e+03 8.791050e+03 8.781853e+03 8.773718e+03
## renovated 2.090816e+04 2.090712e+04 2.090621e+04 2.090542e+04
##
## (Intercept) 6.197996e+06 6.198002e+06 6.198007e+06 6.198011e+06
## bedrooms -3.968972e+04 -3.969013e+04 -3.969049e+04 -3.969081e+04
## bathrooms 4.507413e+04 4.507246e+04 4.507098e+04 4.506966e+04
## sqft_living 1.619410e+02 1.619510e+02 1.619598e+02 1.619676e+02
## sqft_lot 4.782656e-03 4.787630e-03 4.792043e-03 4.795956e-03
## floors 2.741465e+04 2.741640e+04 2.741794e+04 2.741931e+04
## waterfront 6.472892e+05 6.472886e+05 6.472882e+05 6.472877e+05
## view 9.209640e+04 9.209547e+04 9.209464e+04 9.209392e+04
## condition 1.988821e+04 1.988782e+04 1.988748e+04 1.988718e+04
## grade 1.200849e+05 1.200860e+05 1.200870e+05 1.200878e+05
## sqft_above 4.474593e-01 4.372336e-01 4.282009e-01 4.202268e-01
## yr_built -3.575808e+03 -3.575813e+03 -3.575816e+03 -3.575820e+03
## sqft_living15 2.580600e+01 2.580482e+01 2.580378e+01 2.580286e+01
## sqft_lot15 -5.458542e-01 -5.458606e-01 -5.458662e-01 -5.458712e-01
## basement 8.766527e+03 8.760173e+03 8.754563e+03 8.749613e+03
## renovated 2.090472e+04 2.090411e+04 2.090357e+04 2.090310e+04

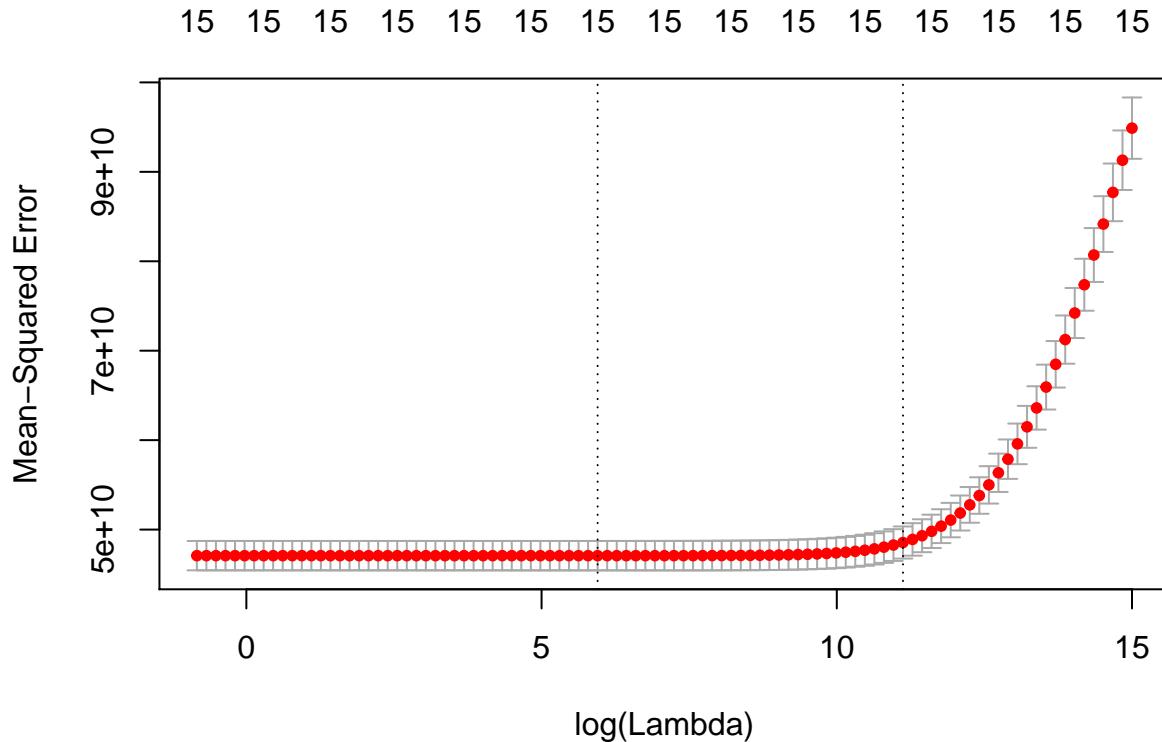
```

```

## 
## (Intercept) 6.198015e+06 6.198018e+06 6.198021e+06 6.198023e+06
## bedrooms -3.969109e+04 -3.969134e+04 -3.969155e+04 -3.969174e+04
## bathrooms 4.506850e+04 4.506746e+04 4.506655e+04 4.506574e+04
## sqft_living 1.619745e+02 1.619806e+02 1.619859e+02 1.619906e+02
## sqft_lot 4.799421e-03 4.802489e-03 4.805202e-03 4.807601e-03
## floors 2.742052e+04 2.742159e+04 2.742253e+04 2.742337e+04
## waterfront 6.472873e+05 6.472870e+05 6.472867e+05 6.472864e+05
## view 9.209328e+04 9.209272e+04 9.209223e+04 9.209180e+04
## condition 1.988692e+04 1.988668e+04 1.988648e+04 1.988630e+04
## grade 1.200886e+05 1.200893e+05 1.200899e+05 1.200904e+05
## sqft_above 4.131911e-01 4.069865e-01 4.015176e-01 3.966995e-01
## yr_built -3.575822e+03 -3.575824e+03 -3.575826e+03 -3.575828e+03
## sqft_living15 2.580205e+01 2.580133e+01 2.580070e+01 2.580015e+01
## sqft_lot15 -5.458756e-01 -5.458795e-01 -5.458830e-01 -5.458860e-01
## basement 8.745247e+03 8.741398e+03 8.738006e+03 8.735019e+03
## renovated 2.090269e+04 2.090233e+04 2.090202e+04 2.090174e+04

# use cross-validation to find the optimal lambda value for the ridge model
set.seed(1)
cv.ridge = cv.glmnet(x, y,
                      alpha = 0,
                      lambda = exp(seq(-1, 15, length = 100)),
                      type.measure = "mse")
par(mfrow = c(1, 1))
plot(cv.ridge)

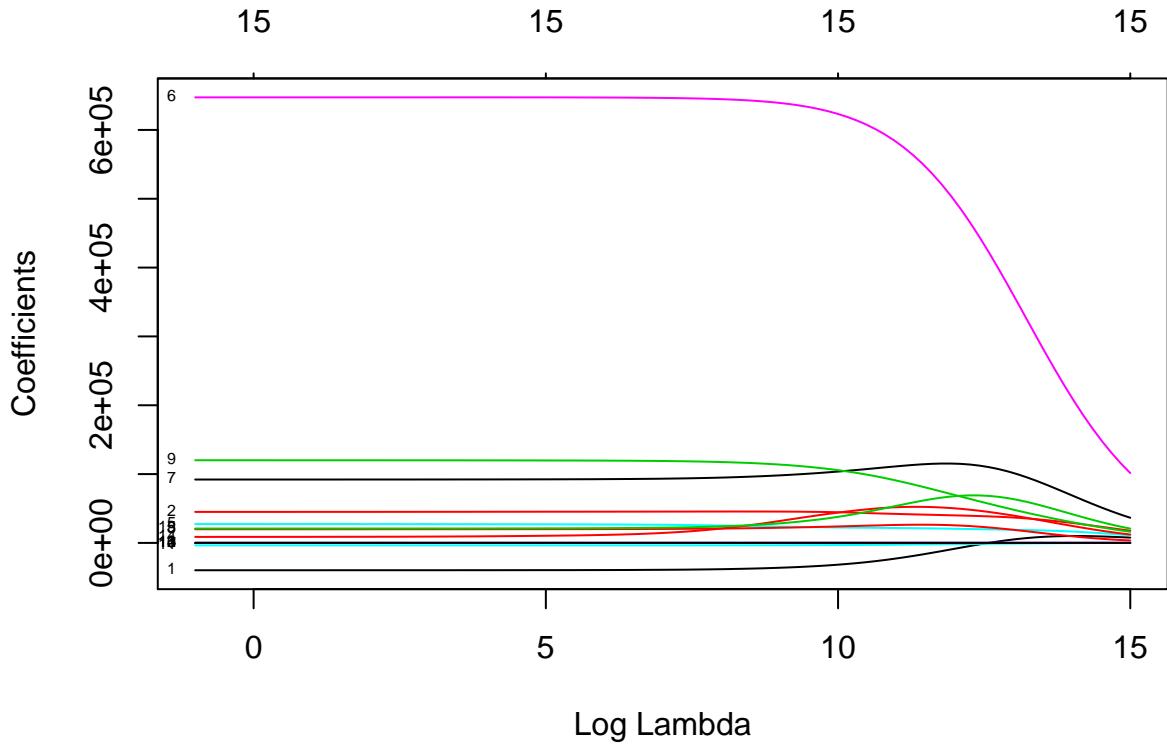
```



```

# generate the trace plot
plot(model_ridge, xvar = "lambda", label = T)

```



```

# obtain the optimal lambda value
best.lambda = cv.ridge$lambda.min; best.lambda

## [1] 383.5596

# obtain the predicted coefficients
predict(model_ridge, s = best.lambda, type = "coefficients")

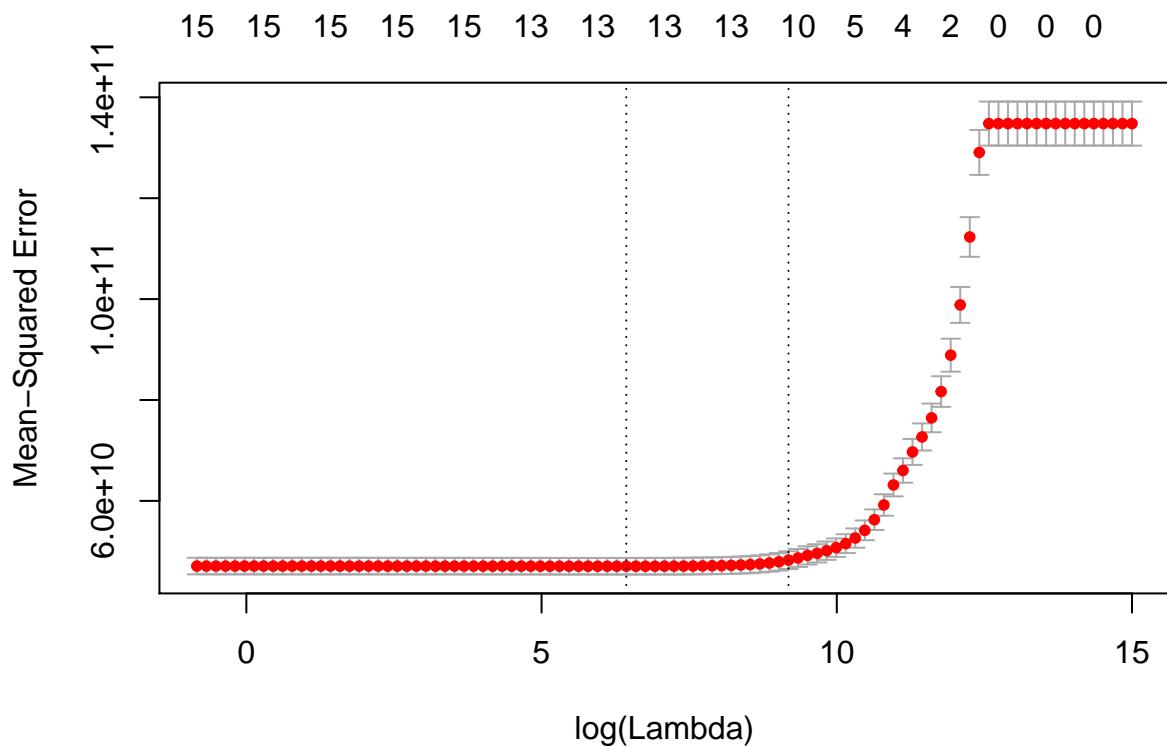
## 16 x 1 sparse Matrix of class "dgCMatrix"
##           1
## (Intercept) 6.181823e+06
## bedrooms    -3.941840e+04
## bathrooms   4.541239e+04
## sqft_living 1.578199e+02
## sqft_lot    3.649836e-03
## floors      2.682397e+04
## waterfront   6.471007e+05
## view        9.263446e+04
## condition   2.009136e+04
## grade       1.195816e+05
## sqft_above   4.743800e+00
## yr_built    -3.567003e+03
## sqft_living15 2.628040e+01
## sqft_lot15   -5.437224e-01
## basement    1.155161e+04
## renovated   2.140516e+04

# use cross-validation to find the optimal lambda value for lasso regression model
set.seed(1)
cv.lasso = cv.glmnet(x, y, alpha = 1, lambda = exp(seq(-1, 15, length = 100)))
cv.lasso$lambda.min

```

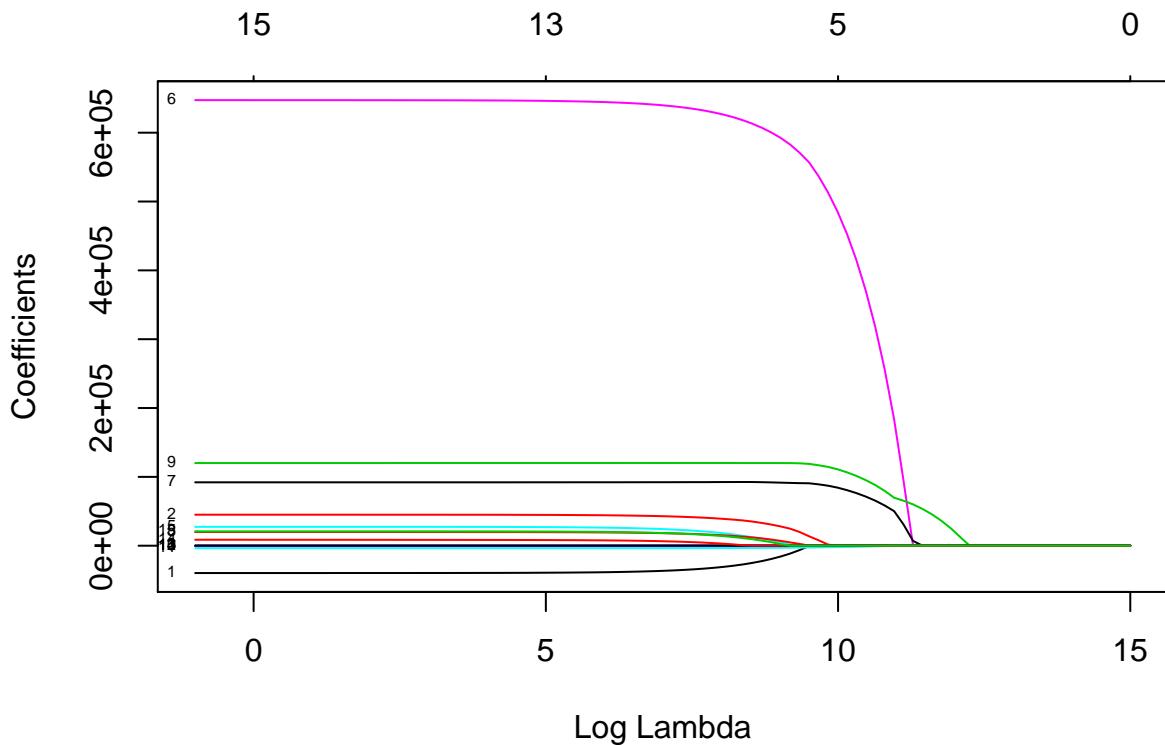
```
## [1] 622.8735
```

```
plot(cv.lasso)
```



```
# plot the result
```

```
plot(cv.lasso$glmnet.fit, xvar = "lambda", label = T)
```



```

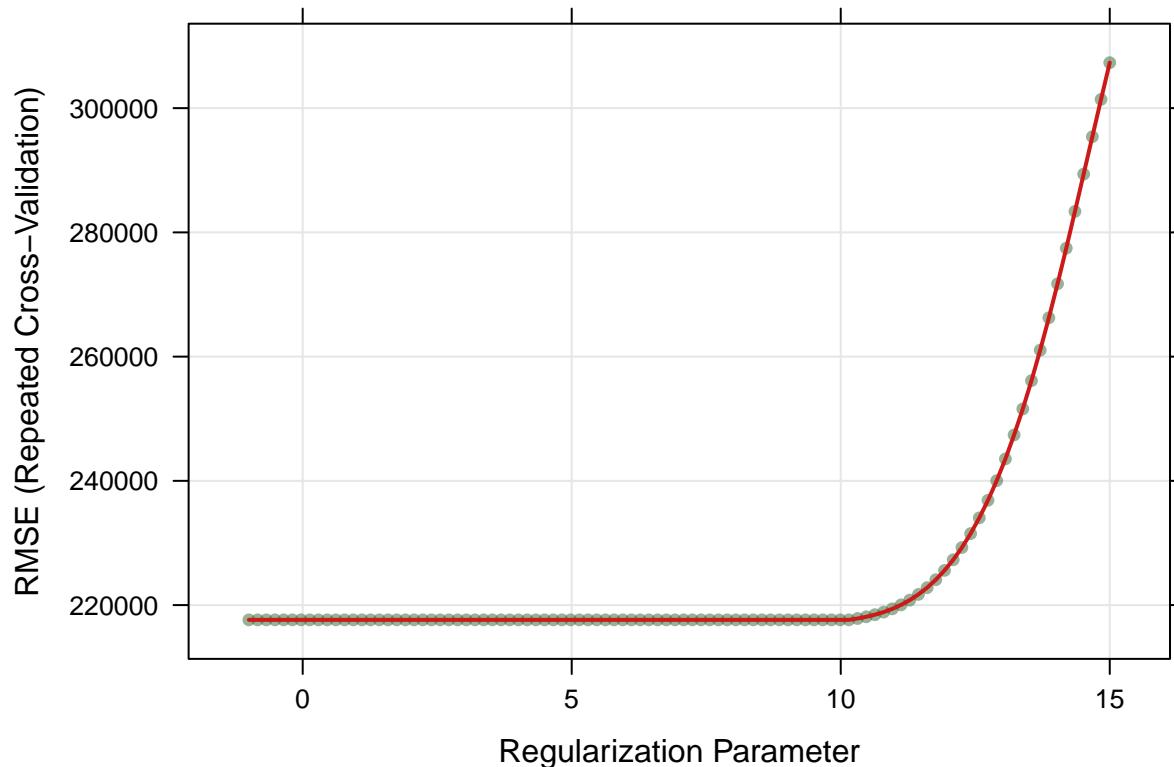
# obtain the predicted coefficients
predict(cv.lasso, s = "lambda.min", type = "coefficients")

## 16 x 1 sparse Matrix of class "dgCMatrix"
##                               1
## (Intercept) 6.117066e+06
## bedrooms     -3.788350e+04
## bathrooms    4.377197e+04
## sqft_living  1.619786e+02
## sqft_lot      .
## floors        2.555757e+04
## waterfront   6.429926e+05
## view          9.219782e+04
## condition    1.892745e+04
## grade         1.201552e+05
## sqft_above    .
## yr_built     -3.531946e+03
## sqft_living15 2.463530e+01
## sqft_lot15   -5.134542e-01
## basement     7.346715e+03
## renovated    1.950752e+04

ctrl1 = trainControl(method = "repeatedcv", number = 10, repeats = 5)
set.seed(849)
# fit a ridge model using caret package
ridge.fit = train(x, y,
                  method = "glmnet",
                  tuneGrid = expand.grid(alpha = 0,
                                         lambda = exp(seq(-1, 15, length = 100))),
                  trControl = ctrl1)

# plot the RMSE by log(lambda)
plot(ridge.fit, xTrans = function(x) log(x))

```



```
# find the optimal lambda
ridge.fit$bestTune

##      alpha    lambda
## 69      0 21805.1
# fit a lasso model using caret
lasso.fit = train(x, y,
                  method = "glmnet",
                  tuneGrid = expand.grid(alpha = 1,
                                         lambda = exp(seq(-1, 15, length = 100))),
                  trControl = ctrl1)

## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info =
## trainInfo, : There were missing values in resampled performance measures.
```