

Historic District Commission Concord, Massachusetts Residential Properties Analysis

Luis D. Berrizbeitia, Vice-Chair HDC

November 2020

Contents

- Introduction 1
- Data sources..... 1
 - Property valuations records..... 1
 - Minimal sale value of a property to be considered in the analysis. 1
 - Date range of transactions included in the analysis 2
 - Historic District designation 2
 - Most recent sale data set..... 3
 - Limitations of the data set 3
 - Proportion of properties in historic districts 3
- Determinants of residential property sale values 4
 - Effect of living area and lot size on sale price 4
 - Effect of Building age 4
 - Effect of the year of the sale..... 5
 - Effect of the age of the building 6
 - Model calibration and diagnostics m_04 7
 - Effect of residential zone 8
 - Additional model diagnostics..... 9
- Relationship between appraised value and sale price 9
 - Data set 9
 - Exploratory analysis 10
 - Model diagnostics 11
- Interim conclusions..... 11
 - Effect of a property being in a historic district 11
- Conclusions 13
- References 13

Historic District Commission Concord, Massachusetts

Residential Properties Analysis

Luis D. Berrizbeitia, Vice-Chair HDC
November 2020

Introduction

An analysis of the factors determining sale and assessed values of historic properties in Concord, Massachusetts was carried out with the purpose of establishing the effect of the property being inside or outside a designated historic district.

Property records supplied by the Town's Assessor and by the Senior Town Planner with the Historic Districts Commission in addition to records available in the Town's web site were used to develop a database with property information. This database was used to formulate and analyze various models and explore the effect of a property being in a historic district in terms of its assessed and sale values. .

The statistical analysis was carried out using the R programming language version 4.0.3 (2020-10-10) running under RStudio Version 1.3.1073. The analysis relied on uni- and multi-variate analysis using the stats::lm procedure. Data with wide spreads were presented as logarithmic transformations as appropriate. The analysis was fully documented in an .Rmd file following the principles of reproducible research [1]. This document is available from the author.

Data sources

Property valuations records

Records of property valuations were obtained from the Board of Assessors, the assessment year reported was 2020. Supplementary data as necessary was obtained from current property records available at the town's website in the database provided by Vision Government Solutions (<http://gis.vgsi.com/concordma/Parcel.aspx?pid=####>) with the last number (####) of the entry being the account number (AcctNum) of the property. The data was supplied in a spreadsheet with the fields described in Table 1- Data fields below.

ReplacementCost (RCN) is not the assessed value of the building, so it does not relate to the assessed (AssessedValue) or appraised (AppraisalValue) value listed. For the purposes of property valuations, the appraised value is a more reliable indicator of the actual value of the property. About 13 properties have vastly differing assessed and appraised values.

The initial data table had **23775** records of which **3593** were duplicates which are discarded. **20182** records became available for analysis. Only Residential properties were considered which includes Zone codes "AA", "A", "B and "C". This left **19153** records. After this step, a unique record number (URC) was added so records could be tracked after additions and deletions.

Minimal sale value of a property to be considered in the analysis.

Many sales are transfers of property without commercial purpose. These can be identified by sale prices of \$1.00 or when the sale price is much lower than the assessed or appraised value. Considering the demographics and urban structure of Concord, only transaction where the sale value was at least \$50,000 were considered in the analysis. After applying this cutoff, **10437** records were available for analysis.

Table 1- Data fields

| Field number | Field name | Field name Concord data | Field number | Field name | Field name Concord data |
|--------------|----------------|--|--------------|-------------------|---|
| 1 | URC | Added variable: Universal record locator | 15 | LivingAreaSqFt | CNS_AREA_LIVING |
| 2 | Location | REM_PRCL_LOCN | 16 | ReplacementCost | CNS_SECT_REPLACE_COST |
| 3 | MAP | REM_MBLU_MAP | 17 | UseCode | REM_USE_CODE |
| 4 | BLOCK | REM_MBLU_BLOCK | 18 | Zone | LND_ZONE |
| 5 | LOT | REM_MBLU_LOT | 19 | Neighborhood | LND_NBHD |
| 6 | UNIT | REM_MBLU_UNIT | 20 | LotSizeSqFt | REM_LEGAL_AREA |
| 7 | AcctNum | REM_ACCT_NUM | 21 | AssessedValueLand | PRC_TTL_APPRAIS_LND |
| 8 | AssessedValue | PRC_TTL_ASSESS | 22 | YearBuiltHDC | Added variable |
| 9 | AppraisalValue | PRC_TTL_APPRAIS | 23 | HDC_Name | Added variable: historic district name |
| 10 | PID | REM_PID | 24 | InHDC | Added variable: in a historic district |
| 11 | SalePrice | SLH_PRICE | 25 | BldAge | Added variable |
| 12 | SaleDate | SLH_SALE_DATE | 26 | PredSalePrice | Added variable: predicted from model m_01 |
| 13 | SaleYear | Added variable | 27 | Difference | Added variable: SalePrice - PredSalePrice |
| 14 | YearBuilt | CNS_AYB | | | |

Date range of transactions included in the analysis

The date range of transactions in the database is from December 31, 1899 to October 20, 2020. For the analysis, only transactions that occurred in 1985 and after were included. This left **8581** records.

Historic District designation

The database described above was updated with a list of properties located in historic districts supplied by the Senior Planner. This added information about whether property was in a historic district and which one. The historic districts identified were:

- Barrett Farm District
- American Mile District
- North Bridge - Monument Square District
- Main Street District
- Hubbardville District
- Church Street District

Most recent sale data set

Given the fact that the assessed value is only relevant to the last transaction and accurate only if the transaction took place in 2020, a dataset with only the most recent transaction for each property was created. The variable `YearsLastSale` captures the years passed since the last transaction.

Limitations of the data set

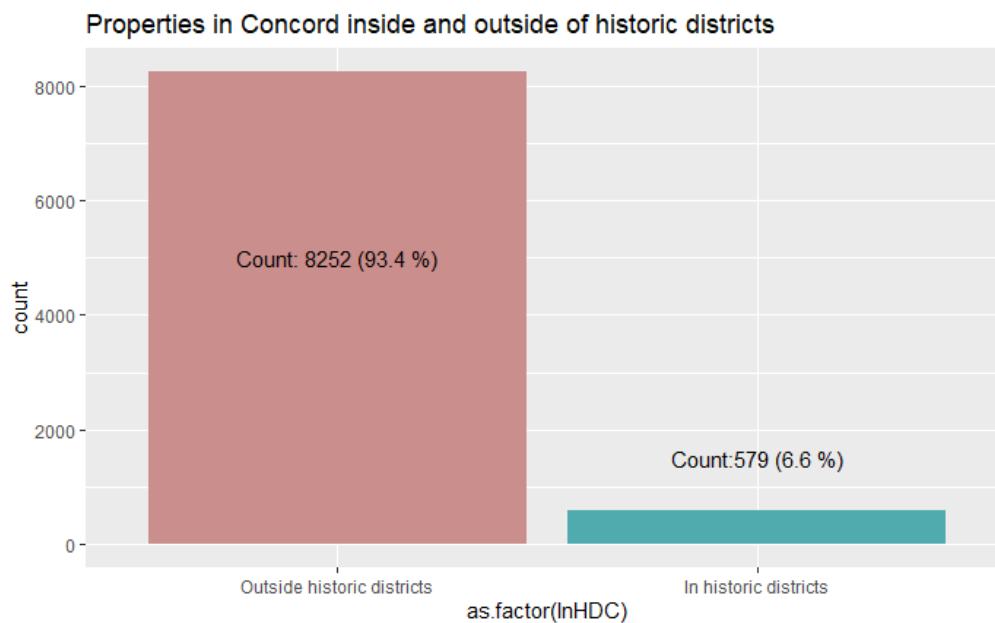
ASSESSED VALUE

The data set supplied by the town's assessor lists all the recorded transactions of a given property which includes the sale price and the date of the transaction. However, **the assessed value listed (variable 8, `AssessedValue`) in each transaction is the last assessed value not the current one at the time of the transaction.** This determines that the assessed value of the property be only relevant if the property was sold in the same year of the assessment which in this case is 2020. This limitation is dealt with a model that estimates the assessed value of the property based on the current (2020) assessed value and the year the transaction took place. In addition, a subset of records is analyzed which only considers the most recent sale and the number of years since that sale.

BUILDING AGE

There are discrepancies for the age of the building between the data supplied by the Historic Districts Commission and the data supplied by the Town's assessor. This discrepancy has a range of -85 years to 313 years with a mean of 10.25 years and a standard deviation of 41.5. Inspection of the data suggests that the records of the Historic District Commission capture the year the principal structure was built while the town's data has the date when major renovations or reconstructions were made. The building age of the town's database is used for the analysis.

Proportion of properties in historic districts



6.55% of properties sold since 1985 are in historic districts. 93.45% are outside historic districts.

Determinants of residential property sale values

The following variables are considered as factors in the sale price of residential properties:

- Living area in sq feet
- Lot size in sq feet
- Building Age
- Year of the sale
- Residential Zone

Effect of living area and lot size on sale price

A model is designed with the property sale price (SalePrice) as the dependent variable and the living area (LivingAreaSqFt) and lot size in sq feet LotSizeSqFt with their interactions as independent variables.

m_01

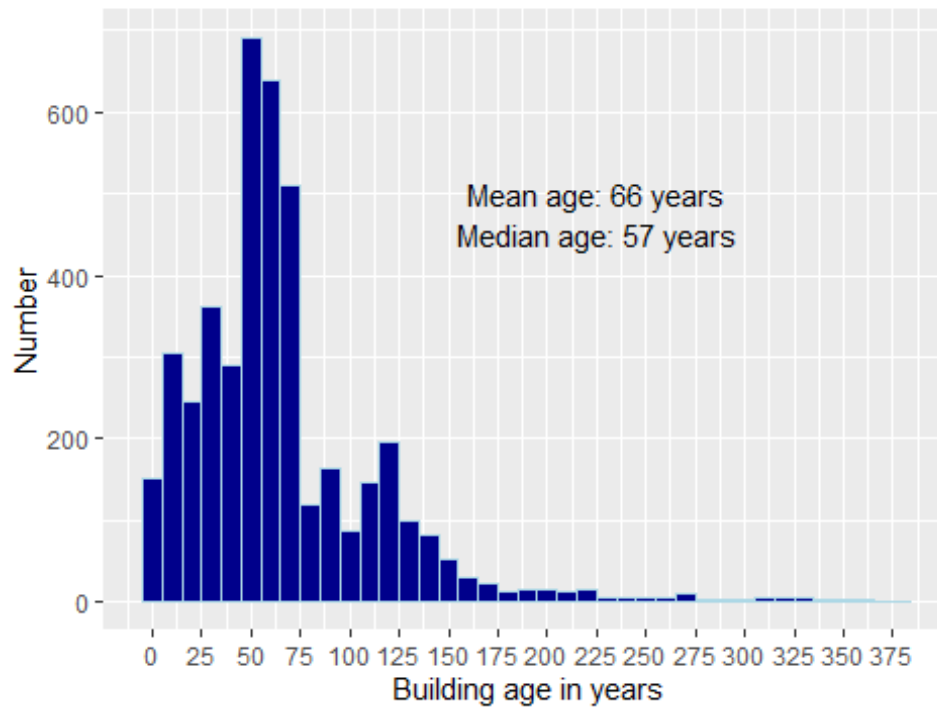
```
## Call:
## lm(formula = SalePrice ~ (LivingAreaSqFt * LotSizeSqFt) - 1,
##     data = Rec2020F, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7662094  -265229  -35584   245663  7956889
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## LivingAreaSqFt      1.727e+02  2.161e+00   79.90  <2e-16 ***
## LotSizeSqFt         4.677e+00  9.301e-02   50.29  <2e-16 ***
## LivingAreaSqFt:LotSizeSqFt -1.813e-04  1.617e-05  -11.21  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 610600 on 8578 degrees of freedom
## Multiple R-squared:  0.7076, Adjusted R-squared:  0.7075
## F-statistic: 6920 on 3 and 8578 DF, p-value: < 2.2e-16
```

The analysis shows living area, lot size, and the year the property was sold are significant components of the sale price. The interactions are significant indicating that the combination of living area and lot size is an additional factor to determine sale price. The model explains 70.8% of the variance between the actual and expected values. The adjusted R-squared is **0.7075089**.

Effect of Building age

Building age is calculated by subtracting the year the property was built from the current year. It is assumed that before 1630, no properties in Concord remain. The average age of properties in residential transactions in Concord is **66** years with a range of **0** to **380** years. 95% of the properties are between **4** and **194** years old.

Age distribution of Concord residential properties



Effect of the year of the sale

The analysis shows the effect of adding the year of the sale (SaleYear) to the model that estimates sale price.

m_02

```
## Call:
## lm(formula = SalePrice ~ (LivingAreaSqFt * LotSizeSqFt * SaleYear) -
##     1, data = Rec2020F, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4542068 -174339  -44883   98051  6525973
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## LivingAreaSqFt    -1.380e+04  3.330e+02  -41.440  <2e-16 ***
## LotSizeSqFt       -2.747e+02  1.215e+01  -22.617  <2e-16 ***
## SaleYear           5.398e+01  6.380e+00   8.461  <2e-16 ***
## LivingAreaSqFt:LotSizeSqFt  3.165e-03  2.437e-03   1.298   0.194
## LivingAreaSqFt:SaleYear     6.974e+00  1.660e-01  42.011  <2e-16 ***
## LotSizeSqFt:SaleYear       1.391e-01  6.047e-03  23.006  <2e-16 ***
## LivingAreaSqFt:LotSizeSqFt:SaleYear -1.690e-06  1.211e-06  -1.395   0.163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 474100 on 8574 degrees of freedom
## Multiple R-squared:  0.8238, Adjusted R-squared:  0.8237
## F-statistic: 5728 on 7 and 8574 DF, p-value: < 2.2e-16
```

Adding the year the property was sold, improves the accuracy of the model. The interactions are significant except those that include the living area combined with the lot size. The model explains 82.4% of the variance between the actual and expected values. The adjusted R-squared is **0.8236801**.

The analysis is rerun without the interaction between living area combined and lot size.

m_03

```
## Call:
## lm(formula = SalePrice ~ ((LivingAreaSqFt + LotSizeSqFt) * SaleYear) -
##     1, data = Rec2020F, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5426615  -174917  -54731    95071   6606885
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## LivingAreaSqFt      -1.422e+04  3.304e+02  -43.05 <2e-16 ***
## LotSizeSqFt         -2.330e+02  7.965e+00  -29.25 <2e-16 ***
## SaleYear              9.041e+01  5.750e+00   15.72 <2e-16 ***
## LivingAreaSqFt:SaleYear  7.174e+00  1.649e-01   43.50 <2e-16 ***
## LotSizeSqFt:SaleYear    1.178e-01  3.962e-03   29.74 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 482000 on 8576 degrees of freedom
## Multiple R-squared:  0.8178, Adjusted R-squared:  0.8177
## F-statistic: 7699 on 5 and 8576 DF,  p-value: < 2.2e-16
```

The model explains 81.8% of the variance between the actual and expected values. The adjusted R-squared is **0.8177011**. Removing the interaction between living area and lot size does not significantly diminish the accuracy of the model.

Effect of the age of the building

Preliminary analysis (not presented here) shows the interaction between living area and building age do not add significant coefficients to the model. The model is formulated without these interactions.

m_04

```
## Call:
## lm(formula = SalePrice ~ (LivingAreaSqFt + LotSizeSqFt + Bld_Age) *
##     SaleYear - 1, data = Rec2020F, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5389400  -178251  -50921    92549   6656074
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## LivingAreaSqFt      -1.328e+04  4.222e+02  -31.457 < 2e-16 ***
## LotSizeSqFt         -2.395e+02  8.510e+00  -28.140 < 2e-16 ***
## Bld_Age              -6.631e+04  1.620e+04  -4.094  4.27e-05 ***
## SaleYear              4.850e+01  7.189e+00   6.747  1.61e-11 ***
## LivingAreaSqFt:SaleYear  6.707e+00  2.107e-01   31.827 < 2e-16 ***
## LotSizeSqFt:SaleYear    1.210e-01  4.231e-03   28.607 < 2e-16 ***
## Bld_Age:SaleYear       3.361e+01  8.088e+00   4.155  3.28e-05 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 475500 on 8565 degrees of freedom
## (9 observations deleted due to missingness)
## Multiple R-squared:  0.822, Adjusted R-squared:  0.8219
## F-statistic: 5652 on 7 and 8565 DF, p-value: < 2.2e-16
```

The model explains 82.2% of the variance between the actual and expected values. The adjusted R-squared is **0.8218997**.

Model calibration and diagnostics m_04

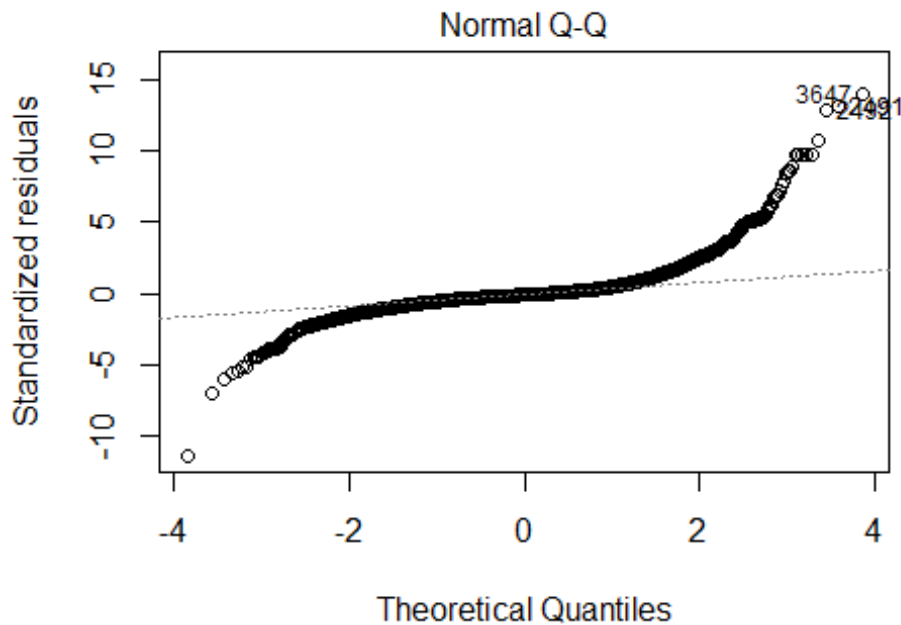
The calibration of the model is analyzed by comparing the predicted and actual sale price of 10 random properties:

Table 2- Predicted and actual sale price of ten random properties.

| | URC | Location | SaleYear | SalePrice | PredSalePrice | Difference |
|----|-------|------------------------|----------|-------------|---------------|-------------|
| 6 | 1337 | 118 BELKNAP ST | 2010 | \$400,000 | \$683,376 | -\$283,376 |
| 2 | 3190 | 1487 MONUMENT ST | 2007 | \$3,500,000 | \$1,244,803 | \$2,255,197 |
| 7 | 15146 | 625 STRAWBERRY HILL RD | 2007 | \$1,177,490 | \$1,063,374 | \$114,116 |
| 5 | 3089 | 147 CAMBRIDGE TPKE | 2003 | \$545,000 | \$510,015 | \$34,985 |
| 3 | 3329 | 15 ROLLINGWOOD LN | 2001 | \$749,000 | \$685,484 | \$63,516 |
| 4 | 1047 | 112 HILL ST | 1998 | \$370,000 | \$372,741 | -\$2,741 |
| 10 | 12482 | 456 BEDFORD ST | 1994 | \$260,000 | \$492,807 | -\$232,807 |
| 8 | 1108 | 113 HUBBARD ST | 1993 | \$360,000 | \$541,441 | -\$181,441 |
| 9 | 12117 | 44 PRAIRIE ST | 1989 | \$243,000 | \$204,827 | \$38,173 |
| 1 | 18662 | 94 ALDEN RD | 1987 | \$263,000 | \$245,030 | \$17,970 |

Diagnostics on the model (Figure 1) show the data is not normally distributed. The Q-Q Plot shows deviations at the extremes of the distribution.

Figure 1- Q-Q plot. Model of Sale Price as a function of (Living Area, Lot Size, Building Age and Sale Year



Effect of residential zone

Different residential zone may influence the sale price of properties. The variable *Zone* is incorporated in the previous model. The reference *Zone* is set to "AA" and the coefficients of the remaining *Zone(s)* are offsets from the reference *ZoneAA*.

m_05

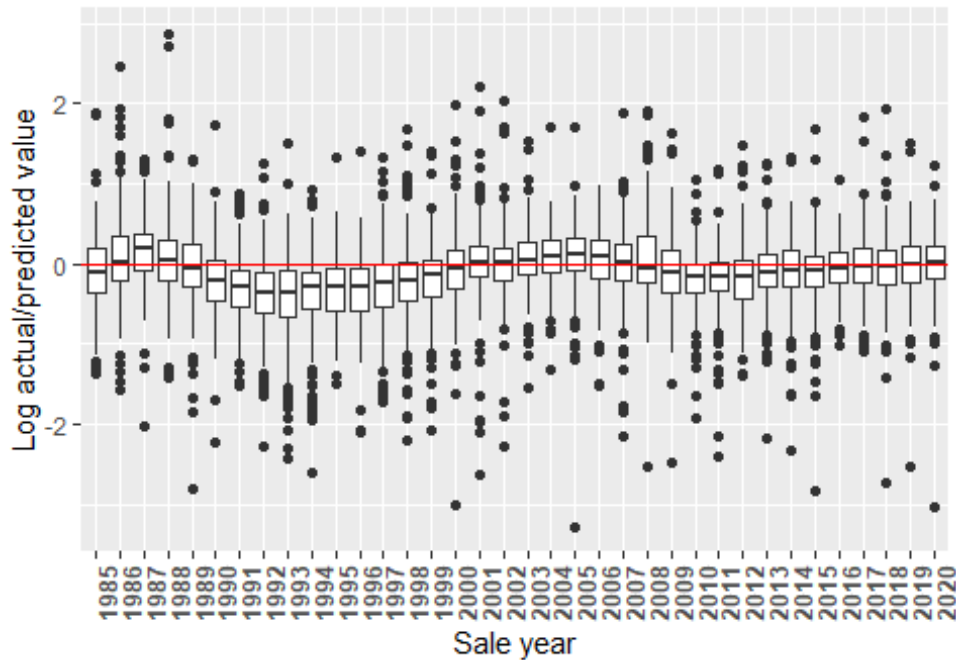
```
## Call:
## lm(formula = SalePrice ~ (LivingAreaSqFt * LotSizeSqFt * SaleYear *
##   Zone) - 1, data = Rec2020F, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4500651 -174602  -36370   94113  6554688
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## LivingAreaSqFt    -9.450e+03  1.235e+03  -7.649  2.24e-14 ***
## LotSizeSqFt      -2.654e+02  1.461e+01 -18.167 < 2e-16 ***
## SaleYear          3.366e+03  2.504e+03   1.345  0.178793
## ZoneAA           -6.591e+06  5.017e+06  -1.314  0.188990
## ZoneA             9.122e+06  6.006e+06   1.519  0.128853
## ZoneB            -1.353e+07  7.049e+06  -1.920  0.054863 .
## ZoneC            -1.026e+07  1.095e+07  -0.937  0.348637
## LivingAreaSqFt:LotSizeSqFt
## LivingAreaSqFt:SaleYear
## LotSizeSqFt:SaleYear
## LivingAreaSqFt:ZoneA
## LivingAreaSqFt:ZoneB
## LivingAreaSqFt:ZoneC
## LotSizeSqFt:ZoneA
## LotSizeSqFt:ZoneB
## LotSizeSqFt:ZoneC
## SaleYear:ZoneA
## SaleYear:ZoneB
## SaleYear:ZoneC
## LivingAreaSqFt:LotSizeSqFt:SaleYear
## LivingAreaSqFt:LotSizeSqFt:ZoneA
## LivingAreaSqFt:LotSizeSqFt:ZoneB
## LivingAreaSqFt:LotSizeSqFt:ZoneC
## LivingAreaSqFt:SaleYear:ZoneA
## LivingAreaSqFt:SaleYear:ZoneB
## LivingAreaSqFt:SaleYear:ZoneC
## LotSizeSqFt:SaleYear:ZoneA
## LotSizeSqFt:SaleYear:ZoneB
## LotSizeSqFt:SaleYear:ZoneC
## LivingAreaSqFt:LotSizeSqFt:SaleYear:ZoneA
## LivingAreaSqFt:LotSizeSqFt:SaleYear:ZoneB
## LivingAreaSqFt:LotSizeSqFt:SaleYear:ZoneC
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 468000 on 8549 degrees of freedom
## Multiple R-squared:  0.8288, Adjusted R-squared:  0.8282
## F-statistic: 1293 on 32 and 8549 DF,  p-value: < 2.2e-16
```

The improvement in the model is minimal. The Adjusted R-squared is **0.8281592** compared with **0.8218997** in the previous model. Adding Zone to the model does not improve its accuracy significantly.

Additional model diagnostics

The difference between predicted and actual sale values is explained by multiple factors, including the skewness as determined by the Q-Q plot above. The following graph shows the differences between sale and predicted values standardized as the ratio of sale value / predicted value by year of sale. Of note is the spread of the data as shown by the many observations outside the 50% quartile.

Figure 2- log[Actual/Predicted] sale value by year. Values above or below zero indicate the sale price was above or below the predicted price.



Relationship between appraised value and sale price

A reliable parameter of the value of a property are the assessed and the appraised valuations used for tax purposes. In Concord, these two quantities are almost always identical. However, as described before, the records do not keep the historic appraised value but use the most recent appraised value in all transactions regardless whether they took place in the same year of the appraisal or several years before. This prevents the use of the appraised value in the calculations of the sale price unless it is adjusted by a given factor. This factor is the years passed since the last transaction which adjusts the appraised value to what it would have been expected at the time of the transaction.

Data set

A data set which only has the most recent transaction for each property is used in this analysis. This data set includes a variable *YearsLastSale* that captures the years passed since the last transaction. The expectation is that transaction that took place in 2020 will have the best correlation with the sale price. As the difference between the year of the appraisal (2020) and the year of the transaction decreases, we can expect a negative coefficient for the appraised value. Under this model, the appraised value would be lower in previous years as a function of the property characteristics and the year of the sale.

Exploratory analysis

The following model explores the relationship between sale price, appraised value, living area and lot size, and the years passed since the last transaction, considering **only the last sale** of the property since 1985.

m_06

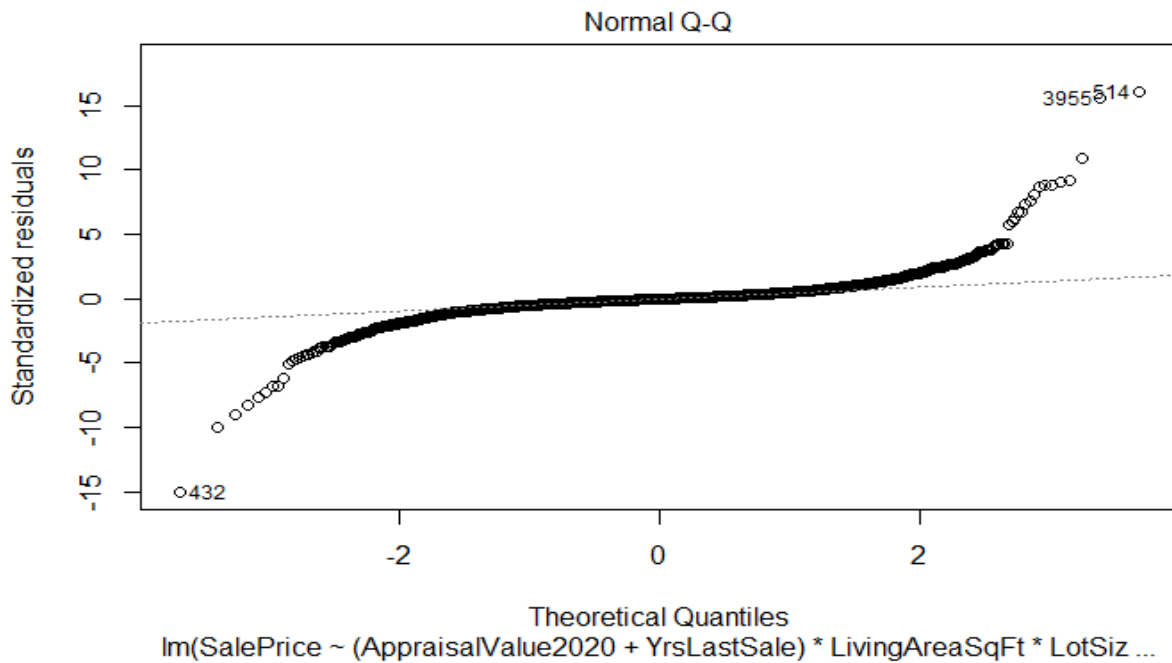
Sale price as a function of Appraisal value, years since last sale, living area and lot size. Only the most recent sale is considered.

```
## Call:
## lm(formula = SalePrice ~ (AppraisalValue2020 + YrsLastSale) *
##   LivingAreaSqFt * LotSizeSqFt - 1, data = Rec2020AP)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3612763  -122304   -9276   107413  5050569
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## AppraisalValue2020    6.019e-01  1.676e-02  35.910 < 2e-16 ***
## YrsLastSale          -3.223e+03  9.297e+02  -3.466 0.000533 ***
## LivingAreaSqFt       1.409e+02  6.547e+00  21.524 < 2e-16 ***
## LotSizeSqFt          5.511e+00  2.002e-01  27.531 < 2e-16 ***
## AppraisalValue2020:LivingAreaSqFt
## -6.967e-06  2.531e-06  -2.753 0.005937 **
## YrsLastSale:LivingAreaSqFt
## -6.163e+00  3.561e-01 -17.308 < 2e-16 ***
## AppraisalValue2020:LotSizeSqFt
## -1.004e-07  2.650e-08  -3.789 0.000153 ***
## YrsLastSale:LotSizeSqFt
## -1.574e-01  5.037e-03 -31.256 < 2e-16 ***
## LivingAreaSqFt:LotSizeSqFt
## -7.848e-04  4.582e-05 -17.126 < 2e-16 ***
## AppraisalValue2020:LivingAreaSqFt:LotSizeSqFt
## 1.038e-10  6.674e-12  15.555 < 2e-16 ***
## YrsLastSale:LivingAreaSqFt:LotSizeSqFt
## 6.014e-06  1.289e-06   4.666 3.16e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 376800 on 4313 degrees of freedom
## Multiple R-squared:  0.9178, Adjusted R-squared:  0.9176
## F-statistic: 4379 on 11 and 4313 DF, p-value: < 2.2e-16
```

The model predicts the sale price well. The adjusted R-squared is 0.9176133. Combining the appraised value in 2020 with the years passed since the past transactions improves the accuracy of the model that predicts the sale price.

Model diagnostics

Figure 3 - Sale price as a function of appraisal, years since last sale, building and lot size.



As in previous models, the model breaks down at the extremes of the distribution.

Interim conclusions

1. Sale value can be modeled using the appraisal value of 2020, the number of years since last transaction, and the building and lot size.
2. The model explains more than 92% of the variability of the data.
3. Regardless of the accuracy of the model, it is not reliable at the extremes of the distribution, that is properties that are very expensive or very inexpensive.

Effect of a property being in a historic district

The property being in a historic district is captured with an indicator variable (InHDC) which is set to zero or one depending whether the property is outside or inside a historic district. The particular historic district is recorded but this data is not used in the analysis. Technically, the interaction between the independent variables identified previously and the indicator variable with a value of zero or one amounts to coefficients that had the variable InHDC are zero for properties outside a historic district and are subsequently discarded. For properties inside historic district the value of the indicator variable is one therefore the coefficient is included in the analysis.

m_06.a

Sale price as a function of Appraisal value, years since last sale, living area, lot size and whether the property is in an historic district. Only the most recent sale is considered.

The model used is similar to m_04 above which has an adjusted R-square of **0.8218997**. The variable InHDC is added as an indicator variable.

```

## Call:
## lm(formula = SalePrice ~ (AppraisalValue2020 + YrsLastSale) *
##   LivingAreaSqFt * LotSizeSqFt * InHDC - 1, data = Rec2020AP)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3605695 -120582   -8401   109075  4991068
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## AppraisalValue2020    5.930e-01  1.772e-02  33.464 < 2e-16 ***
## YrsLastSale          -2.850e+03  9.425e+02  -3.024 0.002506 **
## LivingAreaSqFt       1.409e+02  6.819e+00  20.670 < 2e-16 ***
## LotSizeSqFt          5.559e+00  2.005e-01  27.729 < 2e-16 ***
## InHDC                 1.095e+05  2.198e+05   0.498 0.618452
## AppraisalValue2020:LivingAreaSqFt -7.136e-06  2.568e-06  -2.779 0.005482 **
## YrsLastSale:LivingAreaSqFt      -6.213e+00  3.653e-01 -17.006 < 2e-16 ***
## AppraisalValue2020:LotSizeSqFt   -1.033e-07  2.655e-08  -3.889 0.000102 ***
## YrsLastSale:LotSizeSqFt         -1.577e-01  5.031e-03 -31.335 < 2e-16 ***
## LivingAreaSqFt:LotSizeSqFt      -7.854e-04  4.637e-05 -16.937 < 2e-16 ***
## AppraisalValue2020:InHDC          1.555e-02  1.192e-01   0.130 0.896232
## YrsLastSale:InHDC                 7.684e+02  8.179e+03   0.094 0.925155
## LivingAreaSqFt:InHDC              4.633e+01  6.412e+01   0.723 0.469966
## LotSizeSqFt:InHDC                -7.429e+00  2.754e+00  -2.697 0.007022 **
## AppraisalValue2020:LivingAreaSqFt:LotSizeSqFt 1.039e-10  6.750e-12  15.390 < 2e-16 ***
## YrsLastSale:LivingAreaSqFt:LotSizeSqFt      5.606e-06  1.297e-06   4.323 1.57e-05 ***
## AppraisalValue2020:LivingAreaSqFt:InHDC      3.342e-06  2.775e-05   0.120 0.904152
## YrsLastSale:LivingAreaSqFt:InHDC      -3.419e+00  2.379e+00  -1.437 0.150878
## AppraisalValue2020:LotSizeSqFt:InHDC         1.450e-06  8.942e-07   1.622 0.104901
## YrsLastSale:LotSizeSqFt:InHDC              7.018e-02  8.636e-02   0.813 0.416484
## LivingAreaSqFt:LotSizeSqFt:InHDC           9.271e-04  6.494e-04   1.428 0.153461
## AppraisalValue2020:LivingAreaSqFt:LotSizeSqFt:InHDC -2.368e-10  2.254e-10  -1.051 0.293524
## YrsLastSale:LivingAreaSqFt:LotSizeSqFt:InHDC  1.859e-05  1.842e-05   1.009 0.313012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 374200 on 4301 degrees of freedom
## Multiple R-squared:  0.9192, Adjusted R-squared:  0.9188
## F-statistic: 2127 on 23 and 4301 DF,  p-value: < 2.2e-16

```

The model predicts the sale price well. The adjusted R-squared is 0.9187524. Adding whether the property is inside or outside a historic district does not add any additional information and does not change the expected sale price. None of the coefficients (highlighted in yellow) that include the indicator *InHDC* have coefficients which probability is different from chance. The probabilities of obtaining these coefficients are between 10% and 90%. The only exception is the coefficient that captures the interaction between lot size and being in a historic district (*LotSizeSqFt:InHDC*) which has a probability of occurring by chance of less than 1%. This coefficient is negative indicating that the sale price will decrease as the lot size increases. This may be explained by the fact that in historic districts the houses are older and tend to have smaller lot sizes. Most of the price of the property is driven by the built structures with the size of the lot having a smaller role.

Conclusions

1. The sale value of Concord residential properties can be modeled using the appraisal value of 2020, the number of years since last transaction, and the building and lot size.
2. A model as above explains more than 92% of the variability of the data.
3. None of the models developed is reliable at the extremes of the distribution, that is properties that are either very expensive or very inexpensive.
4. Being in a historic district does not affect the sale price of a property in Concord.

References

1. Gandrud, C., *Reproducible Research with R and R Studio*. 2020: Chapman & Hall.