The Relationship Between Number of Sales and Taxable Value in Property Assessment

Carolyn Logan

6/23/2021

## Abstract

Property tax assessments form the basis of the revenue stream for local and state jurisdictions in Kentucky. These assessments should be levied in a fair and equitable manner by the county tax assessor. Property transfers, or sales, from the open marketplace help determine taxable value, which should indicate fair market value. This paper is interested in determining whether the number of times a property transfers has a statistically significant impact on that property’s tax assessment. This researcher believes that a significant relationship does exist.

Data for this research was provided by a local tax assessor. After perfunctory data cleaning and loading, multiple regression analysis was conducted using a variety of independent variables, including the number of times a property had sold. The multiple regression indicated several times that this variable was statistically significant.

Currently, very little scholarly research exists on this topic. Two similar studies were located and provide interesting insights. These studies together with the results from this study’s regression analysis indicate additional research is warranted. Considering the importance of property tax revenue, additional analysis for low turnover properties would be beneficial for taxing jurisdictions.

## Introduction

In each of Kentucky’s 120 counties, a tax assessor is elected to assess the real estate for tax purposes. According to the Kentucky Constitution, “All property, not exempted from taxation by this Constitution, shall be assessed for taxation at its fair cash value, estimated at the price it would bring at a fair voluntary sale; and any officer, or other person authorized to assess values for taxation, who shall commit any willful error in the performance of his duty, shall be deemed guilty of misfeasance, and upon conviction thereof shall forfeit his office, and be otherwise punished as may be provided by law” (Section 172, n.d.). The tax assessor implements this statutory “fair cash value” requirement in a process called reassessment or revaluation. The assessor analyzes data about properties that have sold. At the same time, the assessor attempts to verify the information on file about the unsold properties. Then, the analysis is used to assign new values to all unsold properties.

The valuation cycle happens every four years, each time with a new set of sales to analyze. Theoretically, after a reassessment occurs, all the properties are assessed at “fair cash value” (section 172, n.d.). Whereas the values computed by the local tax assessor are simply an estimate (Section 172, n.d.), each time a property sells, it is valued accurately by the local real estate market. Properties that rarely sell (low turnover properties) are not exposed to the accuracy of the local real estate market. Instead, they are repeatedly valued with the assessor’s estimate. Subsequently, low turnover properties may not be assessed as accurately as their counterparts with higher turnover.

Since property taxes are directly calculated from property assessments, fair and equitable tax assessments help ensure that the tax burden is distributed evenly. Incorrect property tax assessments cause several problems. First, incorrect assessments lead to inappropriate revenue streams for the taxing jurisdictions. Second, erroneous tax values shift tax burden to accurately assessed properties. Third, flawed tax assessments perpetuate every quadrennial cycle. Finally, incorrect assessments leave the tax assessor open to charges of malfeasance.

This paper will analyze the relationship between the number of times a property has sold and that property’s tax assessment (“value”). Further analysis will be performed using basic characteristics about the homes such as age and size. Based on anecdotal evidence, this researcher believes there is a significant relationship between the number of times a home has sold and its tax value.

## Literature Review

This researcher struggled to locate literature that compared taxable value and the number of times a property experienced turnover. Two similar studies were located that highlight interesting results. Even when a home’s previous price was purposely misstated, that previous price did have an impact on the future market value of the property (Brint, 2009). Although this study was conducted from the buyer’s viewpoint, the tax assessor may be similarly influenced. Considering that the last sale may have occurred decades ago, this bias would result in undervaluing low turnover properties. Additionally, low turnover properties may need repairs and maintenance. The costs to repair a home have significant impact on the home’s selling price (Knight et al, 2000) or market value. This underscores the need for the tax assessor to scrutinize and document the condition of all properties, especially low turnover properties, during the analysis phase of revaluation.

The lack of research on this question and the importance of property assessments points to the need for additional research. This researcher will analyze the relationship between the assessed value and the number of times a property has transferred.

## Research Quesion

Specifically, this study will explore the question:

* Is there a relationship between the number of times a property has sold and its assessed value?

## Theory

This project will use the null hypothesis and alternative hypothesis as stated below.

H0: There is no relationship between the number of times a property has sold and its assessed value.  
H1: A relationship exists between the number of times a property has sold and its assessed value.

The research indicates that a previous sale price has an impact on the next sale price (Brint, 2009). Since assessed value is defined as “fair cash value” (Section 172, n.d.), prior sale price may impact future tax value. Research also suggests that a set of needed home repairs put downward pressure on sales price (Knight et al, 2000). Thus, the fewer times a property has sold may indicate the presence of necessary repairs, which would cause lower prices.

This researcher expects to find a significant relationship between the number of times a property has sold and its tax assessment, based on anecdotal evidence. First, it is commonplace to see these low turnover homes sell for lower values than their higher turnover counterparts. Speculatively, this may be caused by deferred repairs or maintenance. Second, it is not unusual to encounter 30% differences between the assessed values of these two groups of homes, in this researcher’s experience.

## Data

The dataset for this project was furnished by the public assessor’s office. There are three excel datasets: sales, assessed values, and dwelling characteristics. The sales dataset contains sales dates and prices for single-family homes. The assessed values dataset contains recent assessed values. The characteristics dataset includes a set of standard information, including the size of the home and the year the home was built. Each property has been assigned a proprietary key by the assessor’s internal procedures. The files range in size from 80,000 rows to 164,000 rows.

The dwelling characteristic data set included the following information.

## # A tibble: 6 x 17  
## `#` PARID NBHD CDU EXTWALL STORIES STYLE YRBLT GarageSF GarageType SFLA  
## <dbl> <chr> <chr> <chr> <chr> <dbl> <chr> <dbl> <dbl> <chr> <dbl>  
## 1 1 10000~ 187 <NA> 01 2 1 1985 240 ATTACHED 1170  
## 2 2 10000~ 187 <NA> 01 2 1 1984 240 ATTACHED 1170  
## 3 3 10000~ 187 <NA> 01 2 1 1984 240 ATTACHED 1184  
## 4 4 10000~ 187 <NA> 01 1 1 1985 NA <NA> 759  
## 5 5 10000~ 187 <NA> 01 1 1 1985 NA <NA> 759  
## 6 6 10000~ 187 FR 06 1 1 1985 NA <NA> 759  
## # ... with 6 more variables: FIXBATH <dbl>, FIXHALF <dbl>, NumBeds <dbl>,  
## # BSMT <chr>, BsmtSF <dbl>, FBsmtSf <dbl>

The sales data set contained the following elements.

## # A tibble: 6 x 7  
## `#` PARID NBHD SALEDT RECORDDT PRICE SALEKEY  
## <dbl> <chr> <chr> <dttm> <dttm> <dbl> <dbl>  
## 1 536 10000030 187 1998-12-03 00:00:00 NA 72000 488  
## 2 39414 10000030 187 1997-05-20 00:00:00 NA 0 100114  
## 3 94426 10000030 187 2002-07-30 00:00:00 NA 93000 198732  
## 4 538 10000040 187 2000-08-01 00:00:00 NA 75000 161629  
## 5 39415 10000040 187 1987-06-26 00:00:00 NA 64650 100115  
## 6 75734 10000040 187 1989-01-06 00:00:00 NA 64413 489

The assessment data set contained the following information.

## # A tibble: 6 x 3  
## `#` VAL03 PARID   
## <dbl> <dbl> <chr>   
## 1 1 2500 19987980  
## 2 2 3500 16395300  
## 3 3 5000 11448050  
## 4 4 6000 13355650  
## 5 5 6200 10546010  
## 6 6 7500 38288970

First, the data was loaded into a data frame for processing.

## X. PARID NBHD CDU EXTWALL STORIES STYLE YRBLT GarageSF GarageType SFLA  
## 1 1 10000030 187 <NA> 01 2 1 1985 240 ATTACHED 1170  
## 2 2 10000040 187 <NA> 01 2 1 1984 240 ATTACHED 1170  
## 3 3 10000050 187 <NA> 01 2 1 1984 240 ATTACHED 1184  
## 4 4 10000060 187 <NA> 01 1 1 1985 NA <NA> 759  
## 5 5 10000070 187 <NA> 01 1 1 1985 NA <NA> 759  
## 6 6 10000080 187 FR 06 1 1 1985 NA <NA> 759  
## FIXBATH FIXHALF NumBeds BSMT BsmtSF FBsmtSf  
## 1 1 1 NA F 604 326  
## 2 1 1 NA N NA 326  
## 3 1 1 2 N NA 288  
## 4 1 0 NA N NA 0  
## 5 1 0 NA N NA 0  
## 6 1 0 2 N NA 0

Before loading the assessment data set into this data frame, the number of rows for each individual data set was checked. The new data frame contained 81,234 rows while the assessment data set contained 81,154 rows. The merge function was used to merge the assessment data set with the dwelling data frame by joining on the unique key.

## [1] 81234

## [1] 81154

## [1] 81213 19

The resulting data frame contained 81,213 rows and the fields shown below.

## PARID # VAL03 X. NBHD CDU EXTWALL STORIES STYLE YRBLT GarageSF  
## 1 10000030 23311 127500 1 187 <NA> 01 2 1 1985 240  
## 2 10000040 22870 126100 2 187 <NA> 01 2 1 1984 240  
## 3 10000050 24320 131000 3 187 <NA> 01 2 1 1984 240  
## 4 10000060 16886 104900 4 187 <NA> 01 1 1 1985 NA  
## 5 10000070 16887 104900 5 187 <NA> 01 1 1 1985 NA  
## 6 10000080 16913 105000 6 187 FR 06 1 1 1985 NA  
## GarageType SFLA FIXBATH FIXHALF NumBeds BSMT BsmtSF FBsmtSf  
## 1 ATTACHED 1170 1 1 NA F 604 326  
## 2 ATTACHED 1170 1 1 NA N NA 326  
## 3 ATTACHED 1184 1 1 2 N NA 288  
## 4 <NA> 759 1 0 NA N NA 0  
## 5 <NA> 759 1 0 NA N NA 0  
## 6 <NA> 759 1 0 2 N NA 0

The data frame contained columns of data that will not be used in the analysis. These columns are deleted and the edited data frame now contains the following information from the dwelling and assessment data sets.

## PARID VAL03 YRBLT SFLA  
## 1 10000030 127500 1985 1170  
## 2 10000040 126100 1984 1170  
## 3 10000050 131000 1984 1184  
## 4 10000060 104900 1985 759  
## 5 10000070 104900 1985 759  
## 6 10000080 105000 1985 759

## [1] 81213

The sales data set contained multiple transactions for many of the unique properties. Since the analysis of this paper is based on the number of sales, the sales data set is merged with the data frame, joining again on the unique identifier. Then, a count variable is created. This variable counts the number of times a unique identifier occurs in the data set. The final data frame contains the elements shown and the new variable, sales\_counts. There are now 164,457 rows.

#begin counting the number of sales per parid  
nrow(sales) #164,457

## [1] 164457

DF <- merge(merge1, sales, by = 'PARID', all = FALSE) #this merges sales with the rest of data files   
#and puts together with merge1 (dwelling + assessment)  
head(DF)

## PARID VAL03 YRBLT SFLA # NBHD SALEDT RECORDDT PRICE SALEKEY  
## 1 10000030 127500 1985 1170 536 187 1998-12-03 <NA> 72000 488  
## 2 10000030 127500 1985 1170 39414 187 1997-05-20 <NA> 0 100114  
## 3 10000030 127500 1985 1170 94426 187 2002-07-30 <NA> 93000 198732  
## 4 10000040 126100 1984 1170 75734 187 1989-01-06 <NA> 64413 489  
## 5 10000040 126100 1984 1170 538 187 2000-08-01 <NA> 75000 161629  
## 6 10000040 126100 1984 1170 39415 187 1987-06-26 <NA> 64650 100115

nrow(DF)#verifying row count 164,457

## [1] 164587

DF <- add\_count(DF, PARID, name = "sales\_count")   
head(DF)

## PARID VAL03 YRBLT SFLA # NBHD SALEDT RECORDDT PRICE SALEKEY  
## 1 10000030 127500 1985 1170 536 187 1998-12-03 <NA> 72000 488  
## 2 10000030 127500 1985 1170 39414 187 1997-05-20 <NA> 0 100114  
## 3 10000030 127500 1985 1170 94426 187 2002-07-30 <NA> 93000 198732  
## 4 10000040 126100 1984 1170 75734 187 1989-01-06 <NA> 64413 489  
## 5 10000040 126100 1984 1170 538 187 2000-08-01 <NA> 75000 161629  
## 6 10000040 126100 1984 1170 39415 187 1987-06-26 <NA> 64650 100115  
## sales\_count  
## 1 3  
## 2 3  
## 3 3  
## 4 3  
## 5 3  
## 6 3

nrow(DF) #164,457

## [1] 164587

#DF is the DF that contains all my data.

Any columns not needed for the analysis are deleted. The data frame now contains:

## PARID VAL03 YRBLT SFLA PRICE sales\_count  
## 1 10000030 127500 1985 1170 72000 3  
## 2 10000030 127500 1985 1170 0 3  
## 3 10000030 127500 1985 1170 93000 3  
## 4 10000040 126100 1984 1170 64413 3  
## 5 10000040 126100 1984 1170 75000 3  
## 6 10000040 126100 1984 1170 64650 3

## PARID Value YRBLT Area Price sales\_count  
## 1 10000030 127500 1985 1170 72000 3  
## 2 10000030 127500 1985 1170 0 3  
## 3 10000030 127500 1985 1170 93000 3  
## 4 10000040 126100 1984 1170 64413 3  
## 5 10000040 126100 1984 1170 75000 3  
## 6 10000040 126100 1984 1170 64650 3

The data provider requested that the unique identifier be replaced or masked. A randomized string containing ten characters was created and added to the data frame as a new variable named “UniqueKey”. The result is shown below.

#generate random string to use as unique key.   
DF <- DF %>% mutate( UniqueKey = do.call(paste0, Map(stri\_rand\_strings, n=nrow(DF), length=c(5, 4, 1),  
 pattern = c('[A-Z]', '[0-9]', '[A-Z]'))))  
head(DF)

## PARID Value YRBLT Area Price sales\_count UniqueKey  
## 1 10000030 127500 1985 1170 72000 3 MOTXX7293N  
## 2 10000030 127500 1985 1170 0 3 HQGGB2175Y  
## 3 10000030 127500 1985 1170 93000 3 JZARA2635S  
## 4 10000040 126100 1984 1170 64413 3 EITUL7008K  
## 5 10000040 126100 1984 1170 75000 3 WIXIQ4015A  
## 6 10000040 126100 1984 1170 64650 3 OIACK8009K

A new data frame called TheKeys was created to preserve the relationship between the unique key provided with the data and the randomized character string.

## DF.PARID DF.UniqueKey  
## 1 10000030 MOTXX7293N  
## 2 10000030 HQGGB2175Y  
## 3 10000030 JZARA2635S  
## 4 10000040 EITUL7008K  
## 5 10000040 WIXIQ4015A  
## 6 10000040 OIACK8009K

## DF.PARID DF.UniqueKey  
## 164582 94550076 HZIWY0033N  
## 164583 94550076 RAZUZ2652U  
## 164584 96700310 XCQWZ4668Y  
## 164585 96700310 NPLBM9657S  
## 164586 96700310 MJAZN4042T  
## 164587 96701002 EHJDU3242J

## PARID Value YRBLT Area Price sales\_count UniqueKey  
## 1 10000030 127500 1985 1170 72000 3 MOTXX7293N  
## 2 10000030 127500 1985 1170 0 3 HQGGB2175Y  
## 3 10000030 127500 1985 1170 93000 3 JZARA2635S  
## 4 10000040 126100 1984 1170 64413 3 EITUL7008K  
## 5 10000040 126100 1984 1170 75000 3 WIXIQ4015A  
## 6 10000040 126100 1984 1170 64650 3 OIACK8009K

## [1] 164587

Next, multiple occurrences of the provided unique key were deleted from the data frame, which reduced the number of rows to 70,081 unique events.

#Delete the multiple rows based on parid.   
DF <- DF[!duplicated(DF$PARID),] #this deletes duplicate rows that came from sales table  
head(DF)

## PARID Value YRBLT Area Price sales\_count UniqueKey  
## 1 10000030 127500 1985 1170 72000 3 MOTXX7293N  
## 4 10000040 126100 1984 1170 64413 3 EITUL7008K  
## 7 10000050 131000 1984 1184 76900 4 PGJPG5212N  
## 11 10000060 104900 1985 759 45000 2 GDSJX5588E  
## 13 10000070 104900 1985 759 55500 3 MJHFJ4744Z  
## 16 10000080 105000 1985 759 53500 3 JKCFO1427V

nrow(DF) #70,081

## [1] 70081

Per the agreement with the data provider, the unique key “PARID” was deleted from the data set. The result is shown below.

#Final set of transformations  
#Delete PARID  
DF <- select(DF, -PARID)  
head(DF)

## Value YRBLT Area Price sales\_count UniqueKey  
## 1 127500 1985 1170 72000 3 MOTXX7293N  
## 4 126100 1984 1170 64413 3 EITUL7008K  
## 7 131000 1984 1184 76900 4 PGJPG5212N  
## 11 104900 1985 759 45000 2 GDSJX5588E  
## 13 104900 1985 759 55500 3 MJHFJ4744Z  
## 16 105000 1985 759 53500 3 JKCFO1427V

One of the final transformations is to change the assessor’s data element called “YRBLT”, the year the home was built, to age by subtracting YRBLT from 2021. The name of the column was also changed. The final transformed data frame is shown below.

## Value Age Area Price sales\_count UniqueKey  
## 1 127500 36 1170 72000 3 MOTXX7293N  
## 4 126100 37 1170 64413 3 EITUL7008K  
## 7 131000 37 1184 76900 4 PGJPG5212N  
## 11 104900 36 759 45000 2 GDSJX5588E  
## 13 104900 36 759 55500 3 MJHFJ4744Z  
## 16 105000 36 759 53500 3 JKCFO1427V

The descriptive statistics for the final data frame are below.

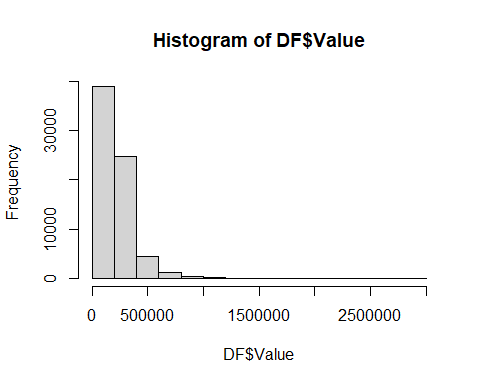
## Value Age Area Price   
## Min. : 2500 Min. : 0.00 Min. : 240 Min. : 0   
## 1st Qu.: 134900 1st Qu.: 26.00 1st Qu.: 1203 1st Qu.: 48500   
## Median : 187500 Median : 43.00 Median : 1525 Median : 103400   
## Mean : 224361 Mean : 46.33 Mean : 1759 Mean : 127375   
## 3rd Qu.: 268000 3rd Qu.: 62.00 3rd Qu.: 2167 3rd Qu.: 170000   
## Max. :3000000 Max. :223.00 Max. :13038 Max. :4053645   
## NA's :8 NA's :100   
## sales\_count UniqueKey   
## Min. : 1.000 Length:70081   
## 1st Qu.: 1.000 Class :character   
## Median : 2.000 Mode :character   
## Mean : 2.349   
## 3rd Qu.: 3.000   
## Max. :10.000   
##

## Methodology

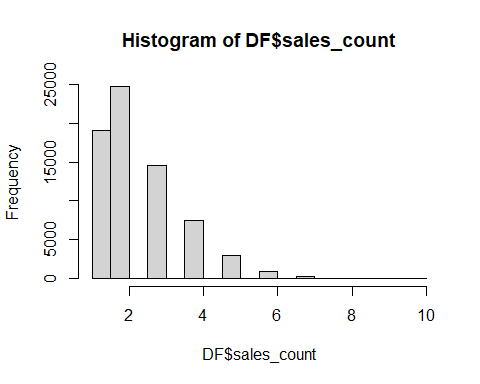
Once the data was cleaned and loaded into one data frame, preparation for analysis began. First, the data structure was reviewed as well as the summary statistics.

## 'data.frame': 70081 obs. of 6 variables:  
## $ Value : num 127500 126100 131000 104900 104900 ...  
## $ Age : num 36 37 37 36 36 36 37 37 37 37 ...  
## $ Area : num 1170 1170 1184 759 759 ...  
## $ Price : num 72000 64413 76900 45000 55500 ...  
## $ sales\_count: int 3 3 4 2 3 3 3 4 3 4 ...  
## $ UniqueKey : chr "MOTXX7293N" "EITUL7008K" "PGJPG5212N" "GDSJX5588E" ...

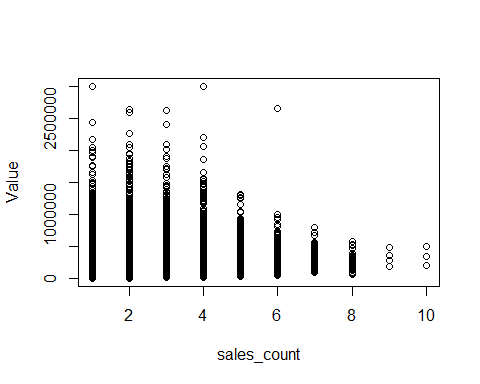
## Value Age Area Price   
## Min. : 2500 Min. : 0.00 Min. : 240 Min. : 0   
## 1st Qu.: 134900 1st Qu.: 26.00 1st Qu.: 1203 1st Qu.: 48500   
## Median : 187500 Median : 43.00 Median : 1525 Median : 103400   
## Mean : 224361 Mean : 46.33 Mean : 1759 Mean : 127375   
## 3rd Qu.: 268000 3rd Qu.: 62.00 3rd Qu.: 2167 3rd Qu.: 170000   
## Max. :3000000 Max. :223.00 Max. :13038 Max. :4053645   
## NA's :8 NA's :100   
## sales\_count UniqueKey   
## Min. : 1.000 Length:70081   
## 1st Qu.: 1.000 Class :character   
## Median : 2.000 Mode :character   
## Mean : 2.349   
## 3rd Qu.: 3.000   
## Max. :10.000   
##

Next, the taxable values were reviewed in a histogram.  


Next, the sales\_count variable was displayed in a histogram.



The sales\_count and taxable values were plotted on a scatter plot.



A pivot table shows the occurrences of each count of sales.

## # A tibble: 10 x 2  
## sales\_count TotalSales  
## <int> <int>  
## 1 1 19108  
## 2 2 49564  
## 3 3 43638  
## 4 4 29956  
## 5 5 14790  
## 6 6 5508  
## 7 7 1589  
## 8 8 368  
## 9 9 36  
## 10 10 30

Several regression models were built and reviewed. The first regression model was created using simple regression. Here, the relationship between the dependent variable taxable value and sales\_count is explored. The summary of the regression coefficients is shown below.

#ran simple regression using sales\_count as independent variable, value as dependent  
simple\_regression <- lm(Value ~ sales\_count, data = DF) #r2 = 0.0015  
summary(simple\_regression)

##   
## Call:  
## lm(formula = Value ~ sales\_count, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -219115 -90305 -37615 43775 2782395   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 212595.7 1251.2 169.9 <2e-16 \*\*\*  
## sales\_count 5009.8 472.8 10.6 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 152600 on 70079 degrees of freedom  
## Multiple R-squared: 0.001599, Adjusted R-squared: 0.001585   
## F-statistic: 112.3 on 1 and 70079 DF, p-value: < 2.2e-16

Since the R-Squared value was insignificant, multiple regression models were built. The first multiple regression analysis uses taxable value as the dependent variable. The number of times a property sold (sales\_count) and the age of the home were the independent variables. The summary is shown below.

#first multiple regression  
#Value is dependent variable, sales\_count + Age are independent variables  
multiple\_regression <- lm(Value ~ sales\_count + Age, data = DF) #r2 = .0.02607  
summary(multiple\_regression)

##   
## Call:  
## lm(formula = Value ~ sales\_count + Age, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -240011 -89306 -40939 38089 2750167   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 256046.29 1612.27 158.81 <2e-16 \*\*\*  
## sales\_count 5089.01 467.03 10.90 <2e-16 \*\*\*  
## Age -941.85 22.45 -41.96 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 150700 on 70078 degrees of freedom  
## Multiple R-squared: 0.02607, Adjusted R-squared: 0.02604   
## F-statistic: 937.8 on 2 and 70078 DF, p-value: < 2.2e-16

The second mulitple regression model again uses taxable value as the dependent variable. The independent variables selected for this analysis were sales\_count and the living area of the home. The summary of this analysis is shown below.

#second multiple regression  
#value is dependent variable, sales\_count + area are independent variables  
MR\_Sales\_Area <- lm(Value ~ sales\_count + Area, data = DF) #r2 = 0.7255  
summary(MR\_Sales\_Area)

##   
## Call:  
## lm(formula = Value ~ sales\_count + Area, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1184064 -40056 -2174 30192 2761556   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.029e+04 9.621e+02 -93.85 <2e-16 \*\*\*  
## sales\_count 8.597e+03 2.477e+02 34.71 <2e-16 \*\*\*  
## Area 1.673e+02 3.893e-01 429.83 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 79870 on 70070 degrees of freedom  
## (8 observations deleted due to missingness)  
## Multiple R-squared: 0.7255, Adjusted R-squared: 0.7255   
## F-statistic: 9.258e+04 on 2 and 70070 DF, p-value: < 2.2e-16

A third and final multiple regression model was built by adding the variable Age to the second model. Again the dependent variable is taxable value. The independent variables are sales\_count, Age, and Area. The summary of this analysis is shown below.

#third multiple regression  
#values is dependent variable, sales\_count + Area + Age are independent variables  
MR\_sales\_Area\_Age <- lm(Value ~ sales\_count + Area + Age, data = DF) #r2 = .7268  
summary(MR\_sales\_Area\_Age)

##   
## Call:  
## lm(formula = Value ~ sales\_count + Area + Age, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1212942 -39944 -1315 31579 2768899   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.048e+05 1.204e+03 -87.06 <2e-16 \*\*\*  
## sales\_count 8.616e+03 2.470e+02 34.89 <2e-16 \*\*\*  
## Area 1.692e+02 3.987e-01 424.22 <2e-16 \*\*\*  
## Age 2.431e+02 1.219e+01 19.95 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 79640 on 70069 degrees of freedom  
## (8 observations deleted due to missingness)  
## Multiple R-squared: 0.727, Adjusted R-squared: 0.727   
## F-statistic: 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16

The chart summarizes the various values of R-Squared calculated by the regression models.

#input the r^2 values  
rSquare <- c(0.001599, 0.02607, 0.7255, 0.727)  
#this will be the titles  
rModel <- c("Simple Regression", "Sales\_Count + Age", "Sales\_count + Area", "Sales\_Count + Age + Area")  
#make a data frame  
RSquareDF <- data.frame("Regression Model" =rModel, "R-Square Value" = rSquare)  
head(RSquareDF)

## Regression.Model R.Square.Value  
## 1 Simple Regression 0.001599  
## 2 Sales\_Count + Age 0.026070  
## 3 Sales\_count + Area 0.725500  
## 4 Sales\_Count + Age + Area 0.727000

## Results

Of the four regression models that were run on this dataset, two indicated a statistically significant relationship between assessed value and number of times a property sold. Both of those models included sales\_count and Area. The R-Squared value for these models was 0.7255 and 0.727, which indicates a strong, positive relationship exists between sales\_count and assessed value.

The coefficients from the regression model indicate that These are summarized in the chart below.

The chart below summarizes the F statistic and its p-value for each of the two multiple regression models with significant R-Squared statistics from this analysis.

## rModel FStatMR3  
## 1 Simple Regression 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16  
## 2 Sales\_Count + Age 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16  
## 3 Sales\_count + Area 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16  
## 4 Sales\_Count + Age + Area 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16  
## FStatMR3.1  
## 1 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16  
## 2 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16  
## 3 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16  
## 4 6.22e+04 on 3 and 70069 DF, p-value: < 2.2e-16

Since the p-values of both F-Statistics are less than 0 (p-value = 2.2e-16), both of these models are statistically significant. Thus, the null hypothesis (“There is no relationship between the number of times a property has sold and its assessed value.”) is rejected in favor of the alternative hypothesis, which stated that there such a relationship does exist.  
## Implications

Given the constitutional mandate for “all property” to be assessed at “fair cash value” (Section 172, n.d.), these results raise questions for further study. Several interesting topics are immediately evident: Do low turnover homes have lower assessed values than medium or high turnover homes? Is there a geographic variable that may be impacting the assessed value for low turnover homes. Can the “drag” on assessed value be estimated or even predicted based on the number of years since the last sale?

Further research should also be conducted on larger data sets. Purposefully, the sales data set used by this study included only one type of sales transaction. In addition, the sales data set was missing several decades (missing dates are prior to 1950) of historical sales.

Finally, further research should be conducted with a geographic variable. Location is one of the primary influencers for a residential property purchaser. Location was not included in this study.

## Conclusion

The purpose of this research was to answer the question, “Is there a relationship between the number of times a property has sold and its tax assessment?”. This researcher theorized that the relationship existed.

The analysis began with three data files from the local tax assessor’s office. The analysis included creating a variable that counted the number of times each property had sold. Subsequently, several multiple regression analyses were performed. The multiple regression R-Squared values indicated a strong, positive relationship between assessed value and the number of times a property had sold. Moreover, the F-statistic and its p-value indicate that both regression models are statistically significant.

# References

Brint, A. (2009). Predicting a house’s selling price through inflating its previous selling price. Journal of the Operational Research Society, 60(3), 339–347. <https://doi.org/10.1057/palgrave.jors.2602567>

Knight, J., Miceli, T., & Sirmans, C. (2000). Repair Expenses, Selling Contracts and House Prices. Journal of Real Estate Research, 20(3), 323–336. <https://doi.org/10.1080/10835547.2000.12091040>

Section 172. (n.d.). <https://apps.legislature.ky.gov/law/constitution>.