

Factors Influencing Housing Prices

Conrad Ibanez

March 2, 2019

####Project Objective

Problem Statement: Selling a house is not likely a familiar process or a common occurrence for many people. However when it does happen, when is the best time to sell? What factors may have an influence for sellers to get top dollar for the homes they are selling? Home ownership is probably one of the biggest investments for many people, so it is important to maximize the price if an owner does decide to sell.

We attempt to answer the following questions:

1. What factors influence the best time to sell?
2. Which time of the year yields the highest prices for homes?
3. Is there a relationship between inventory level and the prices of homes?
4. Is there a specific time such as a particular month where inventory levels are the highest?
5. To get top-dollar, does that mean a homeowner may also need to wait longer to sell?

In this project, housing sales data from Redfin from 2011 to 2018 is analyzed. A subset of the data is selected with a focus on the city of Duluth, GA, with a zip code of 30097. We define the Area 30097 to also include the bordering zip codes of 30096, 30092, 30024, 30022, 30005, and 30043.

```
## [1] "data.frame"
```

```
## 'data.frame': 581 obs. of 58 variables:
## $ i..Worksheet.Filter : chr "Value" "Value" "Value" "Value" ...
## $ Measure.Display : logi NA NA NA NA NA NA ...
## $ Number.of.Records : int 1 1 1 1 1 1 1 1 1 1 ...
## $ Avg.Sale.To.List : num 0.946 0.941 0.955 0.939 0.94 ...
## $ Avg.Sale.To.List.Mom : chr "-0.6%" "-0.2%" "-0.2%" "0.0%" ...
## $ Avg.Sale.To.List.Yoy : chr "-0.9%" "0.3%" "0.8%" "1.6%" ...
## $ City : chr "" "" "" "" ...
## $ Homes.Sold : int 151 192 231 106 69 377 158 124 128 345 ...
## $ Homes.Sold.Mom : chr "-7.9%" "-3.5%" "-15.1%" "-0.9%" ...
## $ Homes.Sold.Yoy : chr "8.6%" "11.0%" "30.5%" "45.2%" ...
## $ Inventory : int 441 533 606 234 217 420 177 228 229 478 ...
## $ Inventory.Mom : chr "5.5%" "-7.1%" "-5.0%" "-0.8%" ...
## $ Inventory.Yoy : chr "-24.1%" "-32.9%" "-28.5%" "-32.6%" ...
## $ Median.Dom : num 116 114 126 169 113 ...
## $ Median.Dom.Mom : num -15.5 -13 -17 0 10 -0.5 3 -0.5 -4 -1 ...
## $ Median.Dom.Yoy : num -0.5 -19.5 10 53 -26 -19 -36.5 23.5 -7 -41.5 ...
## $ Median.List.Ppsf : num 98.1 103.2 79.8 97 98.7 ...
## $ Median.List.Ppsf.Mom : num 0.01052 0.00589 -0.00998 0.07174 0 ...
```

```

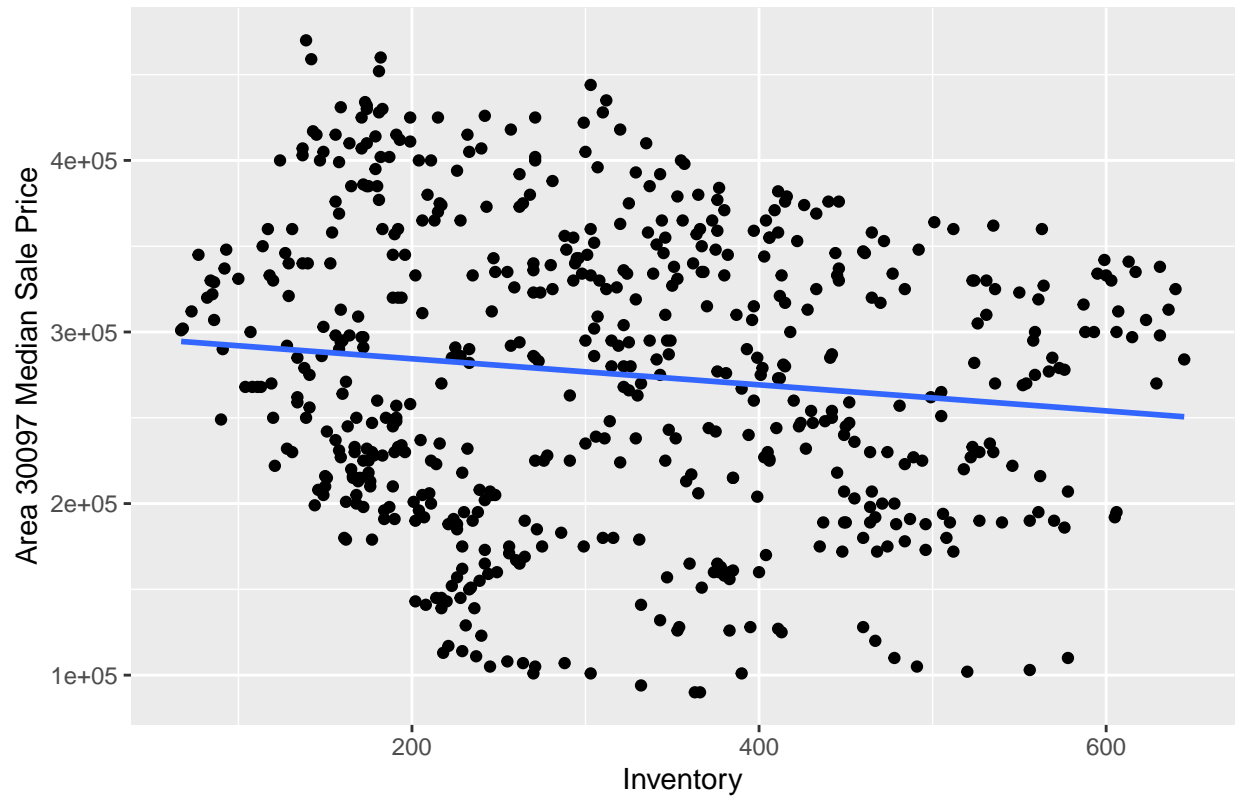
## $ Median.List.Ppsf.Yoy      : num  0.00155 -0.04613 -0.01077 0.31451 -0.04456 ...
## $ Median.List.Price         : num  291376 225000 219900 234900 299900 ...
## $ Median.List.Price.Mom     : num  -0.0284 -0.0385 0.0134 0.0931 0.0169 ...
## $ Median.List.Price.Yoy     : num  -0.0773 -0.0621 0 0.3241 0.0451 ...
## $ Median.Ppsf               : num  93 95.6 76.2 67.8 98.2 ...
## $ Median.Ppsf.Mom           : num  -0.02034 -0.01128 -0.00283 0.0842 -0.0017 ...
## $ Median.Ppsf.Yoy           : num  0.00142 -0.04856 -0.02101 -0.04712 -0.0457 ...
## $ Median.Sale.Price         : num  285000 235000 195000 151000 270000 260000 179000 286000 218000 ...
## $ Median.Sale.Price.Mom     : chr  "-4.6%" "-6.3%" "-2.5%" "8.6%" ...
## $ Median.Sale.Price.Yoy     : chr  "0.0%" "-2.0%" "-12.4%" "4.1%" ...
## $ months_of_supply          : num  NA NA NA NA NA NA NA NA NA NA ...
## $ months_of_supply_mom      : num  NA NA NA NA NA NA NA NA NA NA ...
## $ months_of_supply_yoy      : num  NA NA NA NA NA NA NA NA NA NA ...
## $ New.Listings              : int  190 193 235 99 85 253 117 170 167 361 ...
## $ New.Listings.Mom          : chr  "19.5%" "-14.2%" "-2.1%" "-9.2%" ...
## $ New.Listings.Yoy          : chr  "2.2%" "-15.0%" "-7.1%" "-9.2%" ...
## $ off_market_in_two_weeks   : num  0.0684 0.0466 0.0681 0.0404 0.0588 ...
## $ off_market_in_two_weeks_mom : num  -0.01334 0.00663 0.00142 -0.00547 -0.00271 ...
## $ off_market_in_two_weeks_yoy : num  0.03079 0.01139 0.03251 0.02206 0.00674 ...
## $ pending_sales             : int  149 187 264 94 93 241 110 115 103 340 ...
## $ pending_sales_mom         : num  -0.1337 -0.0209 -0.0222 -0.1296 0.0333 ...
## $ pending_sales_yoy         : num  -0.04487 0.00538 0.28155 0.20513 -0.02105 ...
## $ Period.Begin              : Date, format: "2011-11-01" "2011-11-01" ...
## $ Period.Duration           : int  90 90 90 90 90 90 90 90 90 90 ...
## $ Period.End                 : Date, format: "2012-01-31" "2012-01-31" ...
## $ Price.Drops               : num  NA NA NA NA NA NA NA NA NA NA ...
## $ Price.Drops.Mom           : num  NA NA NA NA NA NA NA NA NA NA ...
## $ Price.Drops.Yoy           : num  NA NA NA NA NA NA NA NA NA NA ...
## $ Property.Type             : chr  "All Residential" "All Residential" "All Residential" "All Residential" ...
## $ Region                    : chr  "Zip Code: 30097" "Zip Code: 30022" "Zip Code: 30024" "Zip Code: 30024" ...
## $ Region.Type               : chr  "zip code" "zip code" "zip code" "zip code" ...
## $ Sold.Above.List           : num  0.1391 0.0938 0.1558 0.0755 0.0435 ...
## $ Sold.Above.List.Mom       : num  -0.01946 -0.00173 0.00879 -0.01799 -0.02062 ...
## $ Sold.Above.List.Yoy       : num  -0.0552 0.0359 0.0372 -0.0204 -0.0389 ...
## $ State                     : chr  "Georgia" "Georgia" "Georgia" "Georgia" ...
## $ State.Code                : chr  "GA" "GA" "GA" "GA" ...
## $ Table.Id                  : int  12459 12391 12393 12454 12375 12391 12454 12375 12454 12411 ...
## $ BeginMonthNum             : chr  "11" "11" "11" "12" ...
## $ BeginMonth                : chr  "Nov" "Nov" "Nov" "Dec" ...
## $ BeginYear                 : chr  "2011" "2011" "2011" "2011" ...

```

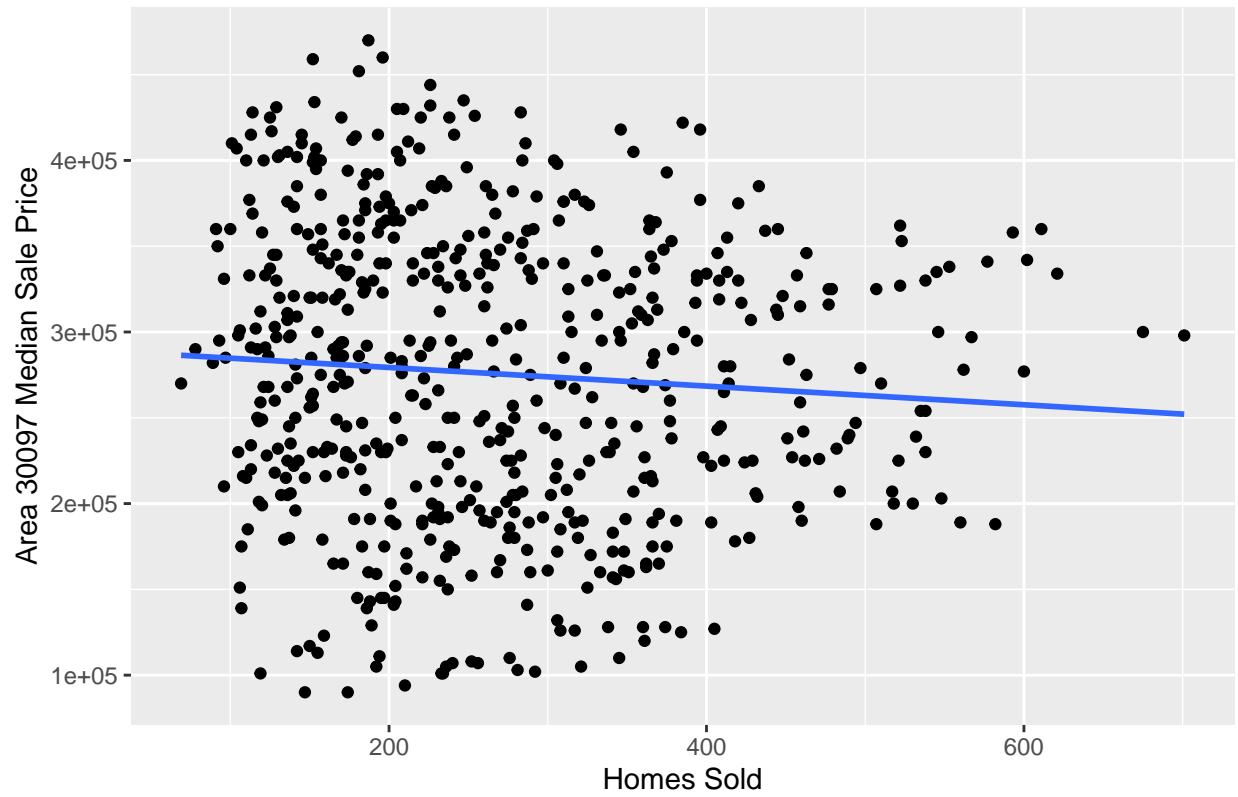
Linear Regression Line on Scatterplots of Median Sale Price and Other Variables

There is a negative relationship between Median Sale Price with most other variables such as Inventory, Homes Sold, New Listings, and Pending Sales. That is when these variables increase, Median Sale Price decreases. However, there is a strong positive relationship between Median Sale Price and Median List Price. This makes sense in that as Median List Price increases, then Median Sale Price also increases.

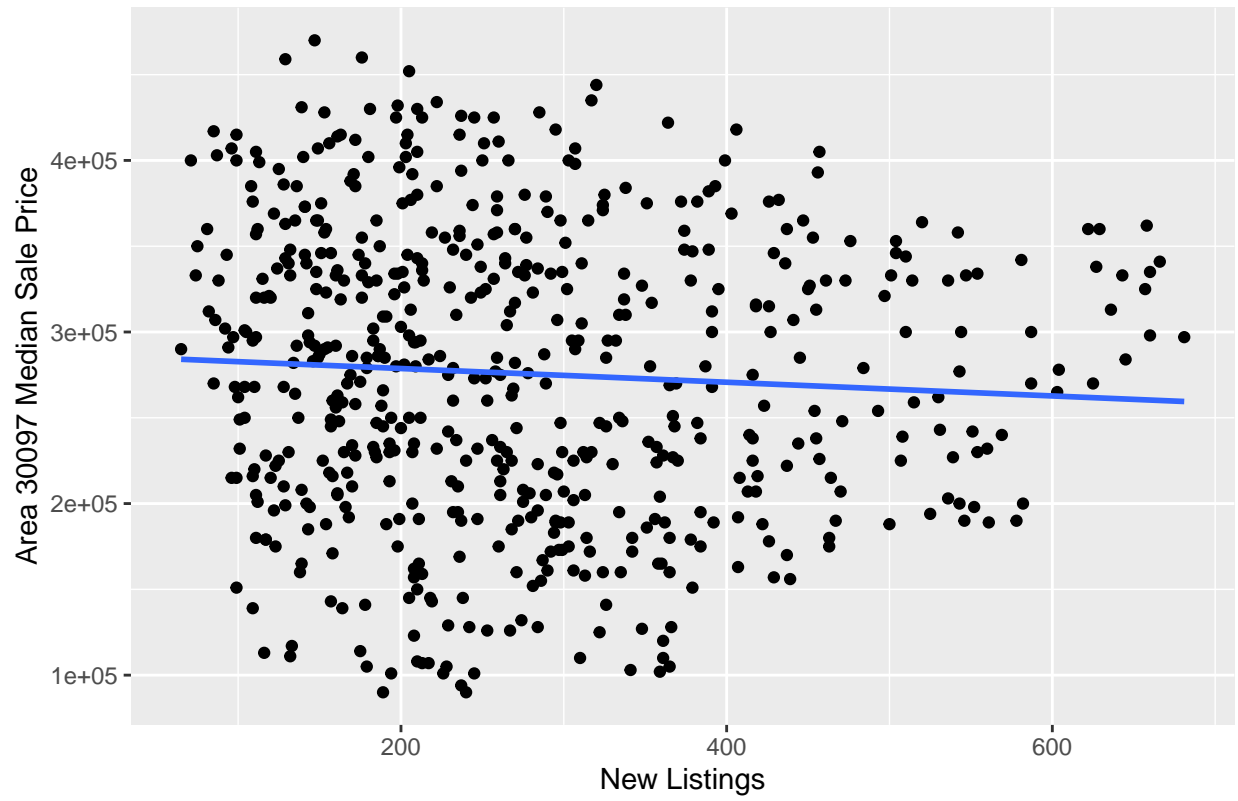
Area 30097 Median Sale Price vs Inventory



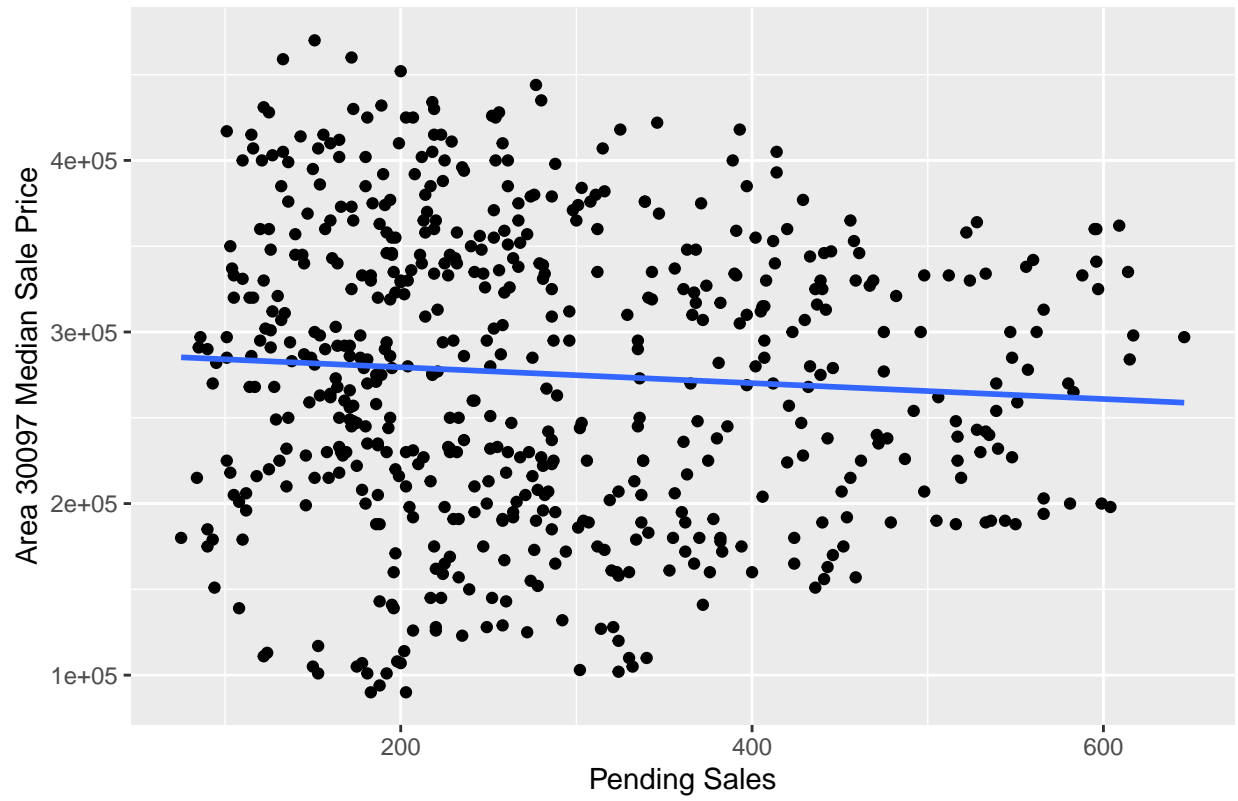
Area 30097 Median Sale Price vs Homes Sold



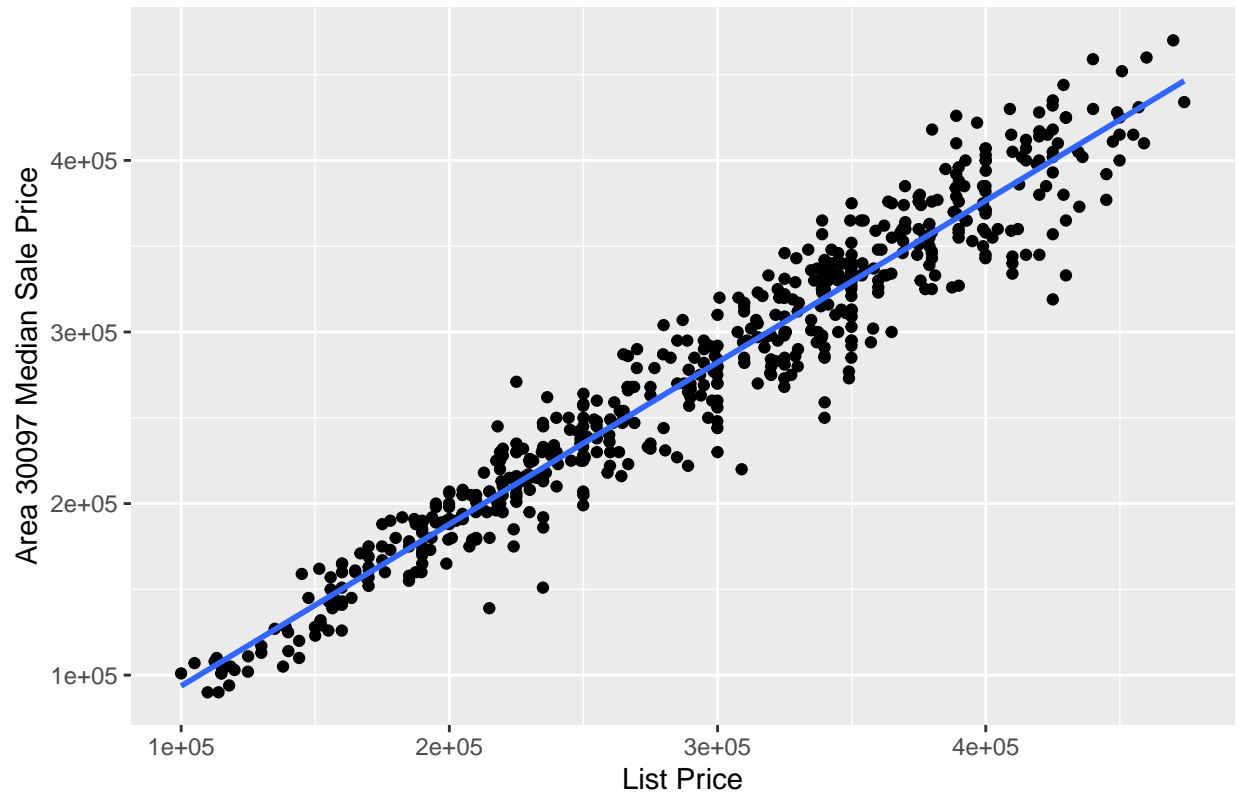
Area 30097 Median Sale Price vs New Listings



Area 30097 Median Sale Price vs Pending Sales



Area 30097 Median Sale Price vs Median List Price

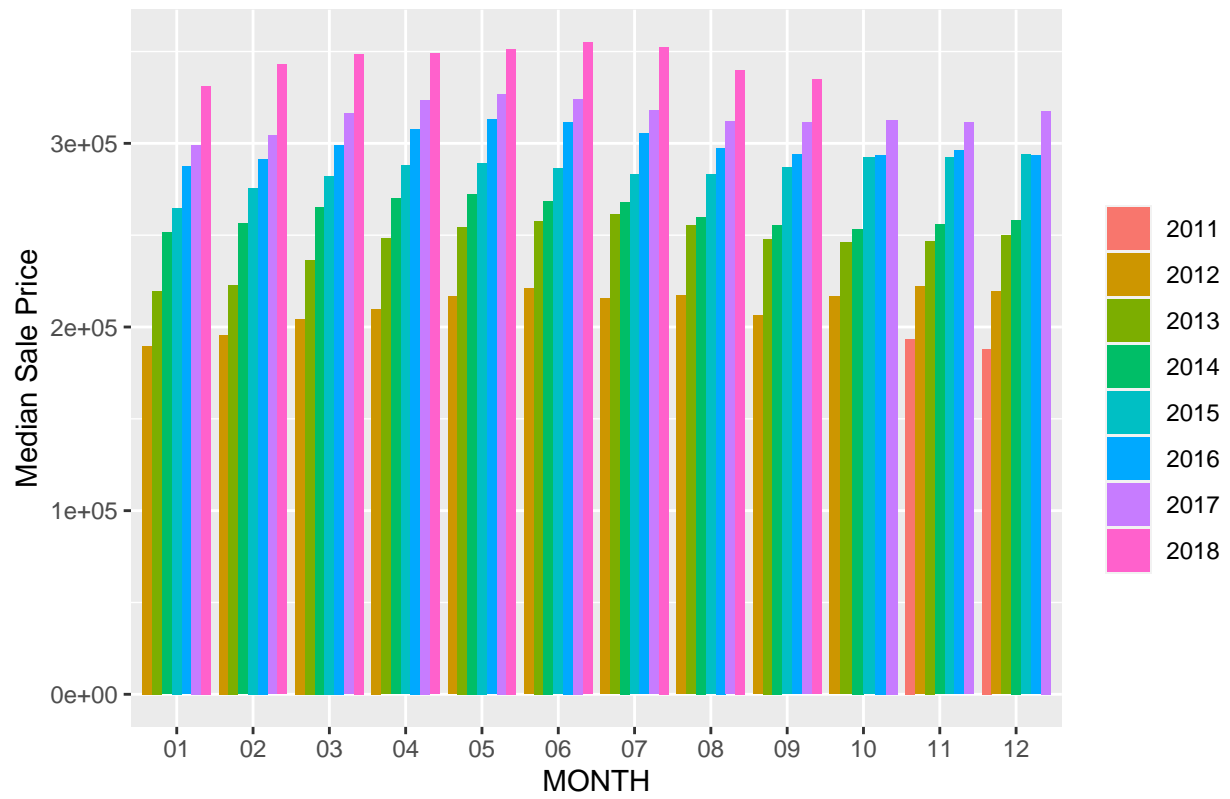


####We have created Time Series Graphs below for the analysis of the variables.

####Median Sale Price Analysis

The Median Sale Price is highest in the month of June, followed by May, July and then April. The Median Sale Price is lowest in the month of November, followed by December and January.

Median Sale Price By Month



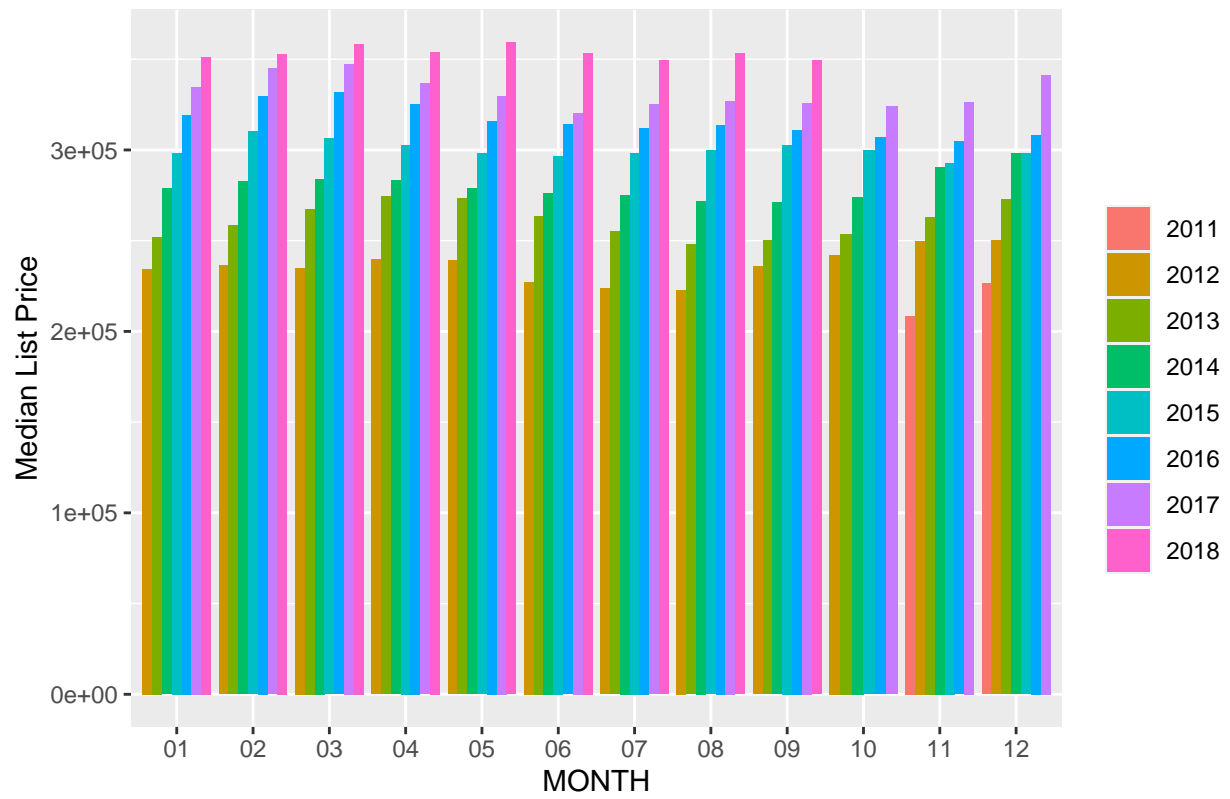


```
## # A tibble: 12 x 2
##   month Median.Sale.Price
##   <chr>          <dbl>
## 1 01          263327.
## 2 02          269898.
## 3 03          278918.
## 4 04          285204.
## 5 05          289061.
## 6 06          289204.
## 7 07          286245.
## 8 08          280653.
## 9 09          276653.
## 10 10         269119.
## 11 11         259694.
## 12 12         260041.
```

####Median List Price Analysis

Median List Price is highest in March, followed by April and February. It is lowest in November, followed by October and December.

Median List Price By Month

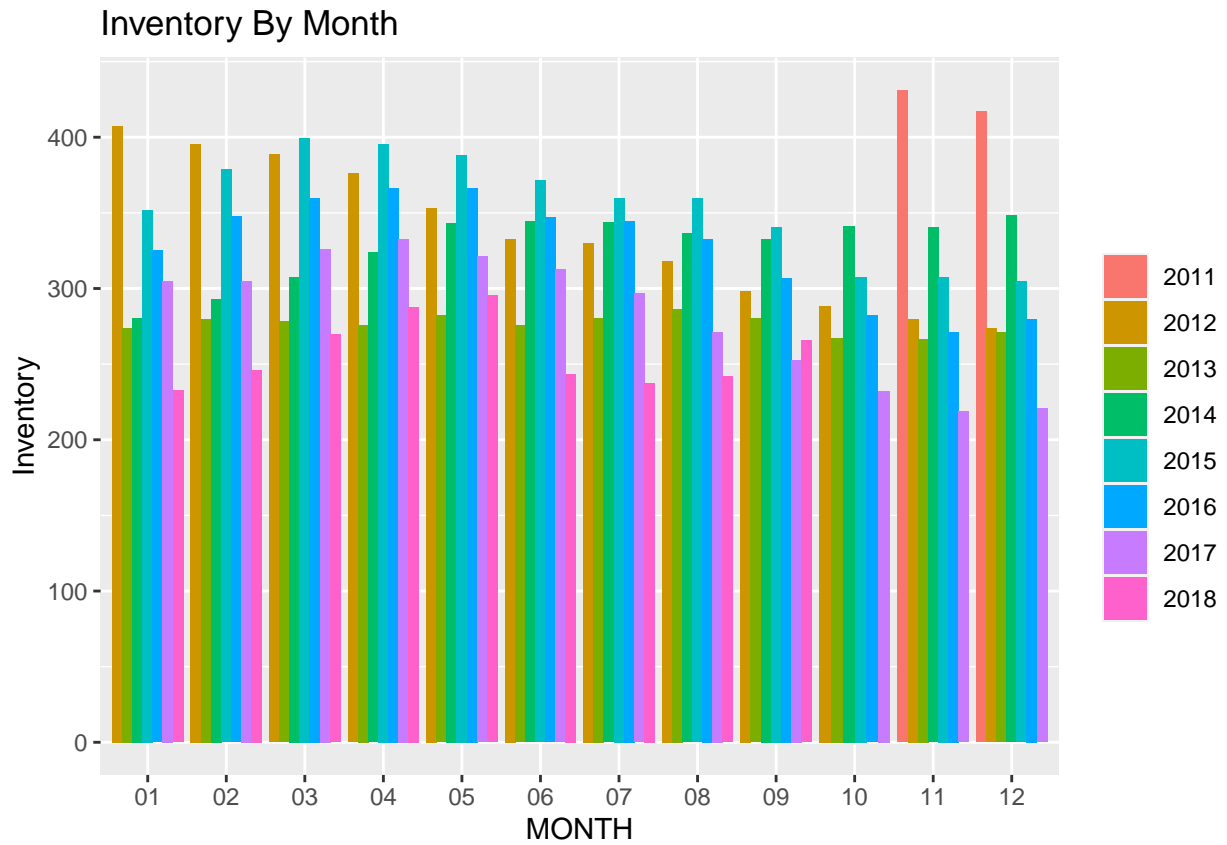


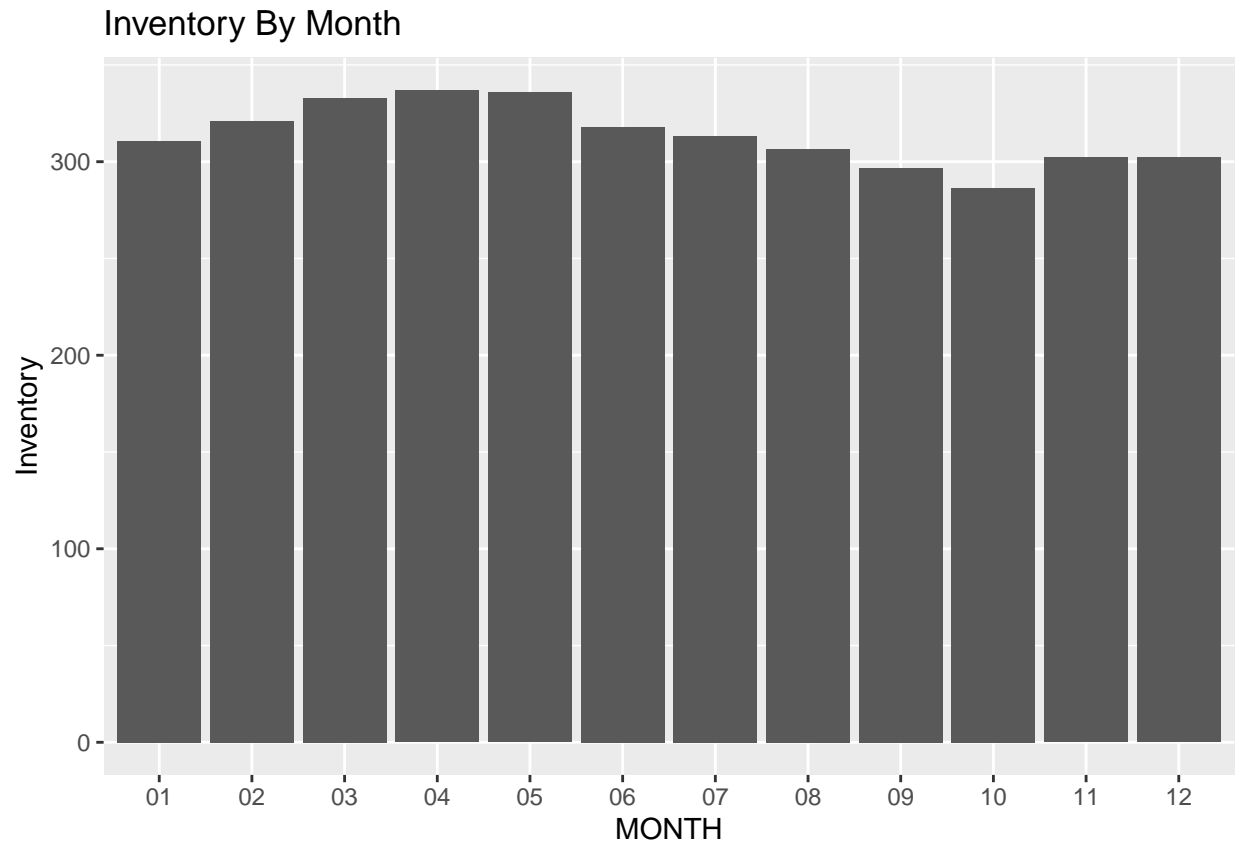


```
## # A tibble: 12 x 2
##   month Median.List.Pricee
##   <chr>          <dbl>
## 1 01          295501.
## 2 02          302148.
## 3 03          304259.
## 4 04          302272.
## 5 05          299160.
## 6 06          293068.
## 7 07          291211.
## 8 08          290829.
## 9 09          292238.
## 10 10         283381.
## 11 11         276550.
## 12 12         284962.
```

####Inventory Analysis

Inventory is at the highest in April, followed by May and March. Inventory is at the lowest in October.

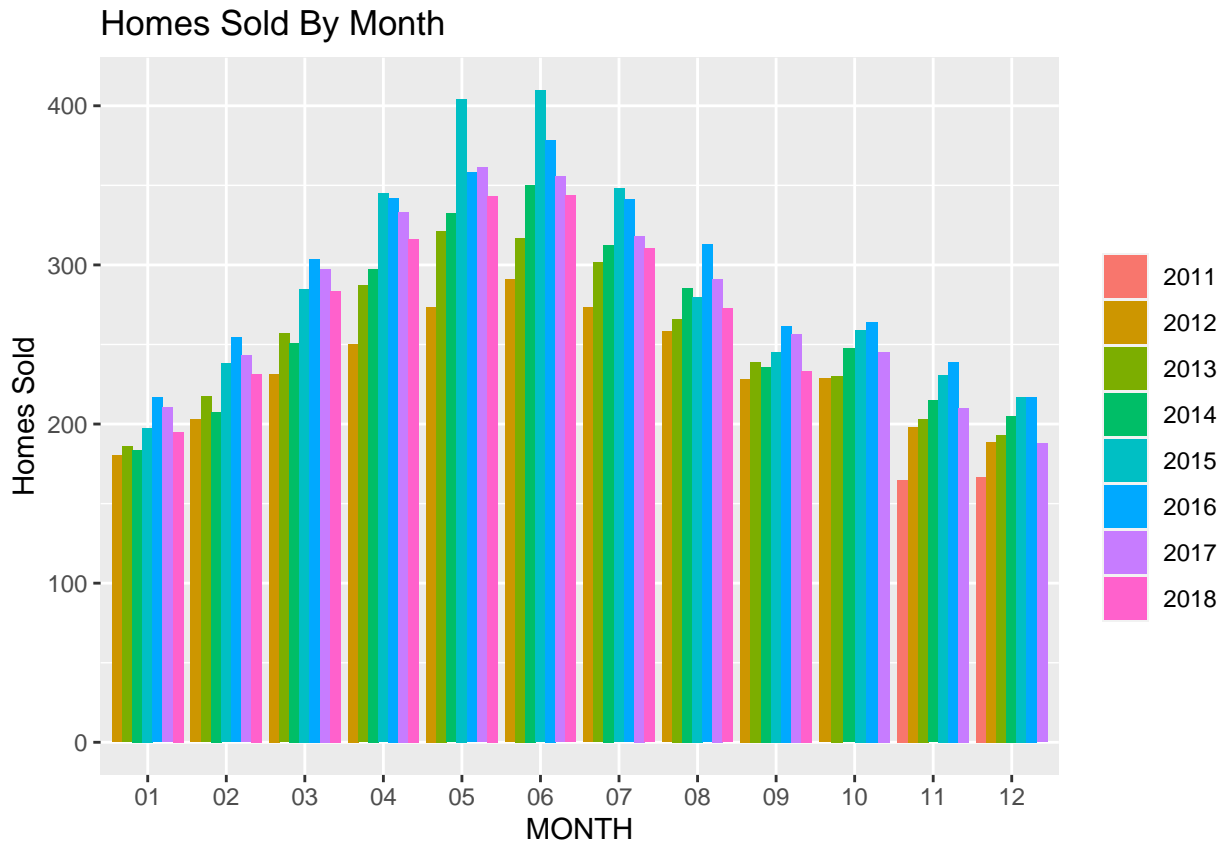




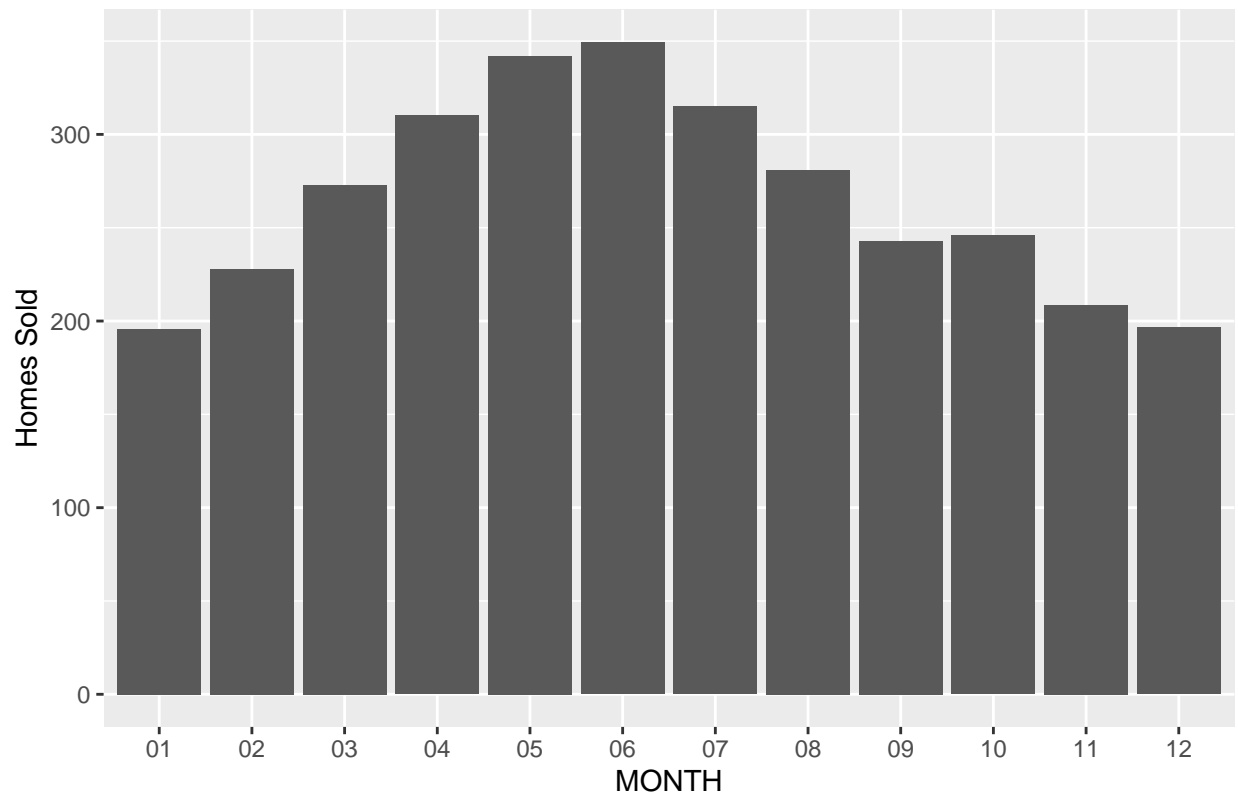
```
## # A tibble: 12 x 2
##   month Inventory
##   <chr>      <dbl>
## 1 01         311.
## 2 02         321.
## 3 03         333.
## 4 04         337.
## 5 05         336.
## 6 06         318.
## 7 07         313.
## 8 08         307.
## 9 09         297.
## 10 10        286.
## 11 11        302.
## 12 12        302.
```

####Homes Sold Analysis

Homes Sold is highest in the month of June, followed by May, July and then April. Homes sold is lowest in the month of January, followed by December and November.



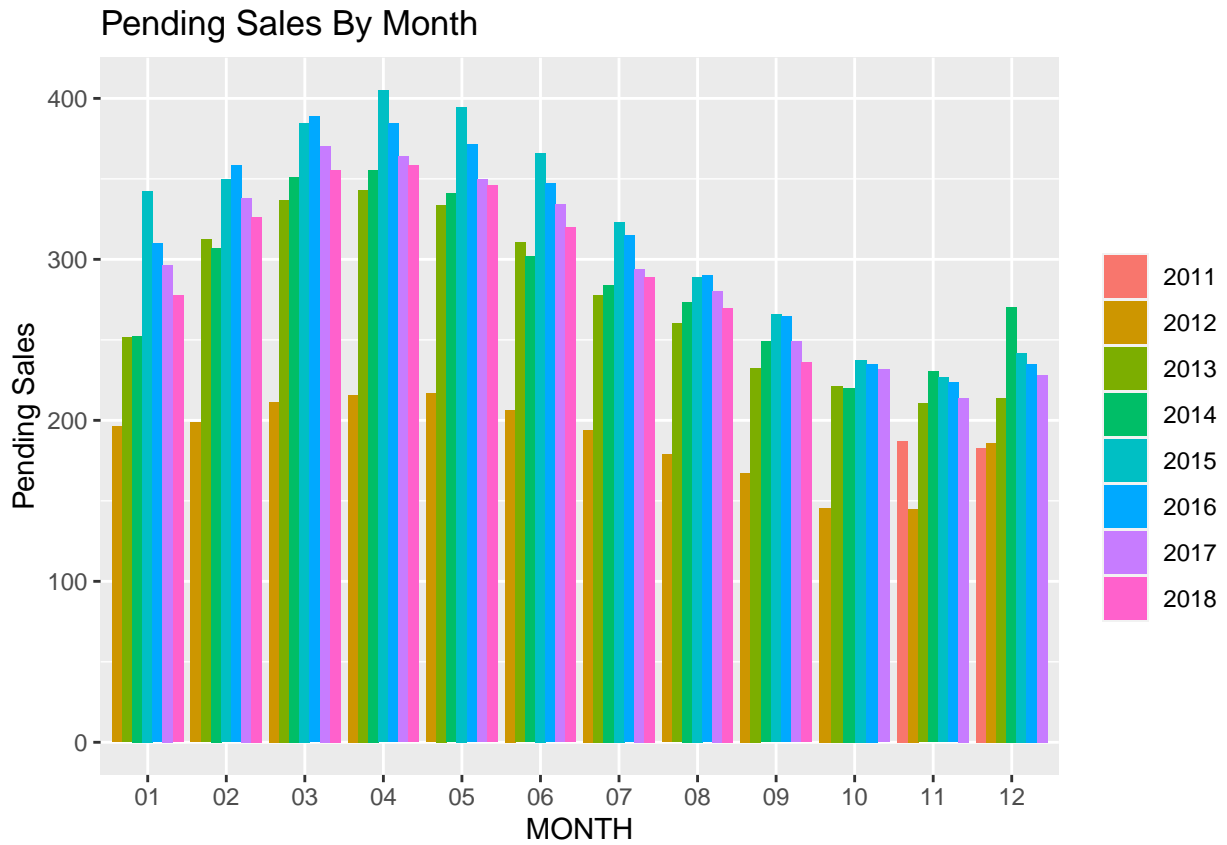
Homes Sold By Month



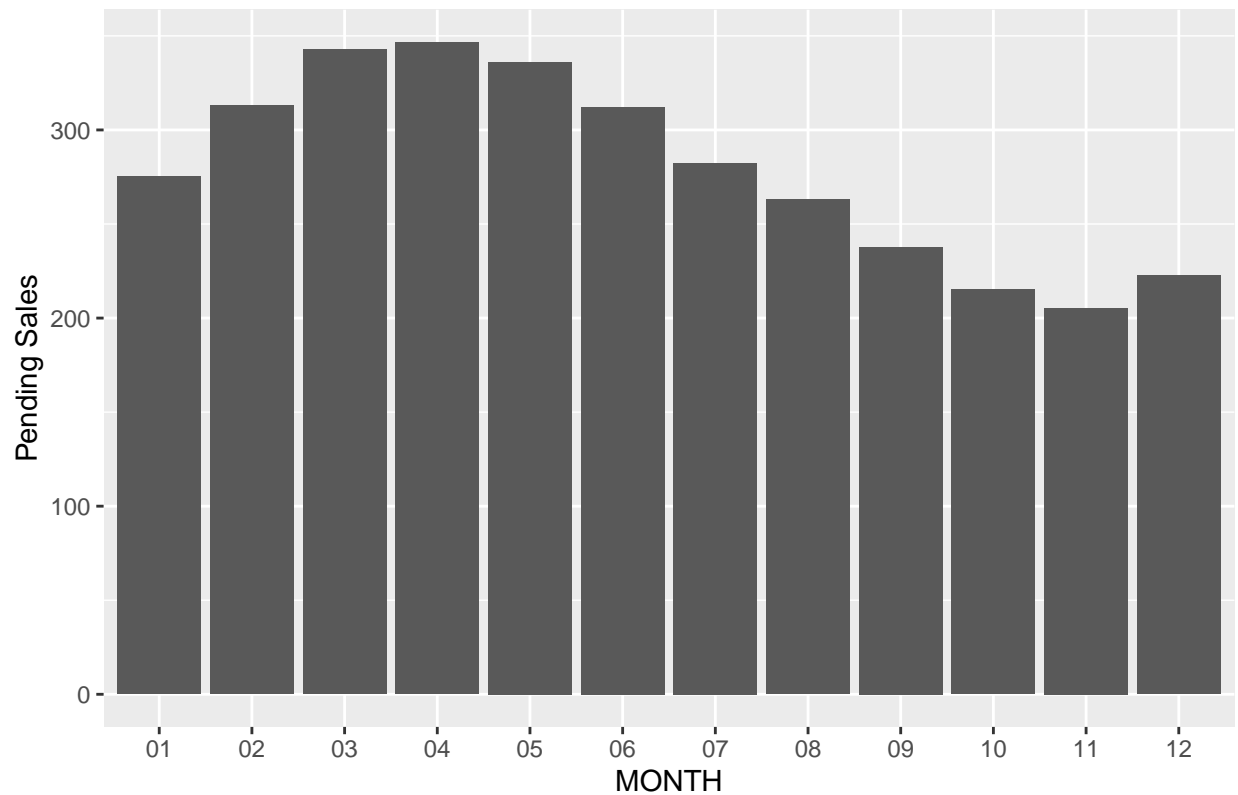
```
## # A tibble: 12 x 2
##   month Homes.Sold
##   <chr>      <dbl>
## 1 01         196.
## 2 02         228.
## 3 03         273.
## 4 04         310.
## 5 05         342.
## 6 06         349.
## 7 07         315.
## 8 08         281.
## 9 09         243.
## 10 10        246.
## 11 11        209.
## 12 12        197.
```

Pending Sales Analysis

Pending Sales is at the highest in April, followed by March and May. Pending Sales is at the lowest in November.



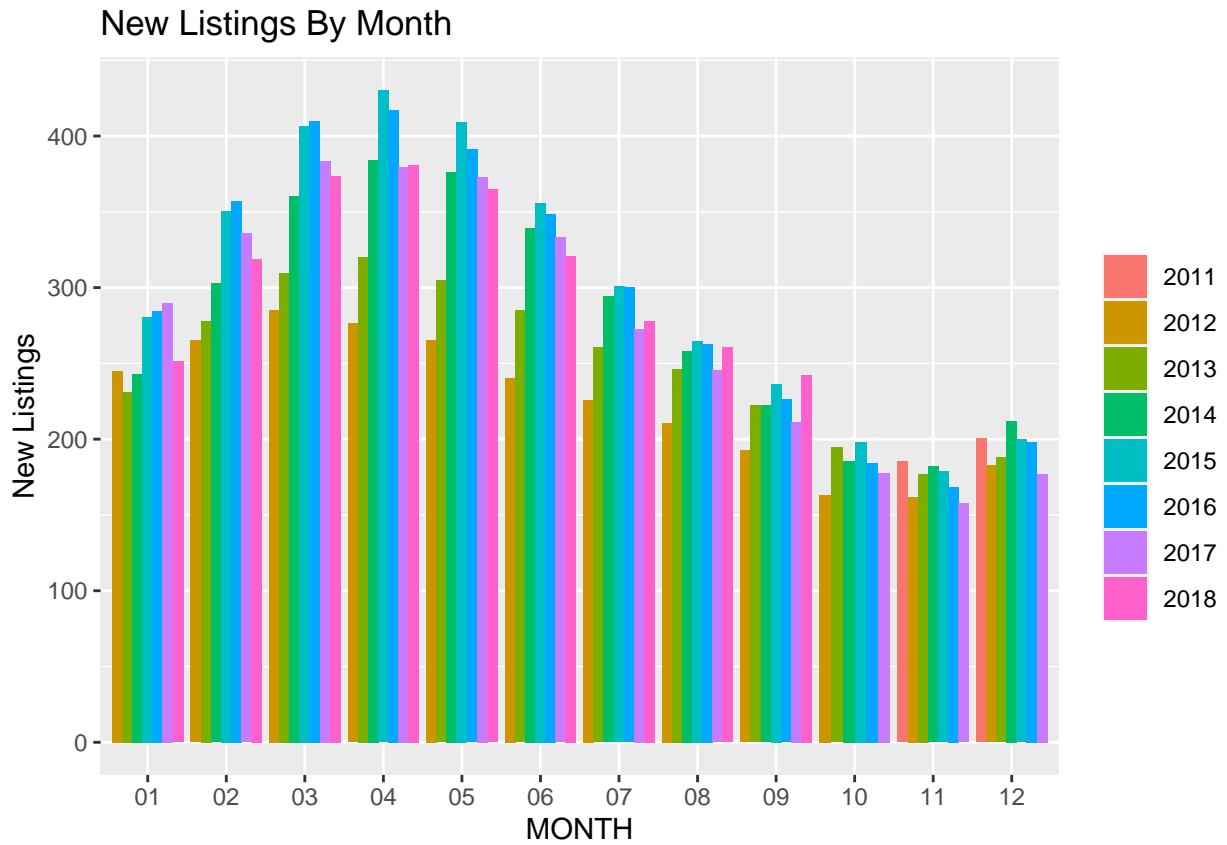
Pending Sales By Month

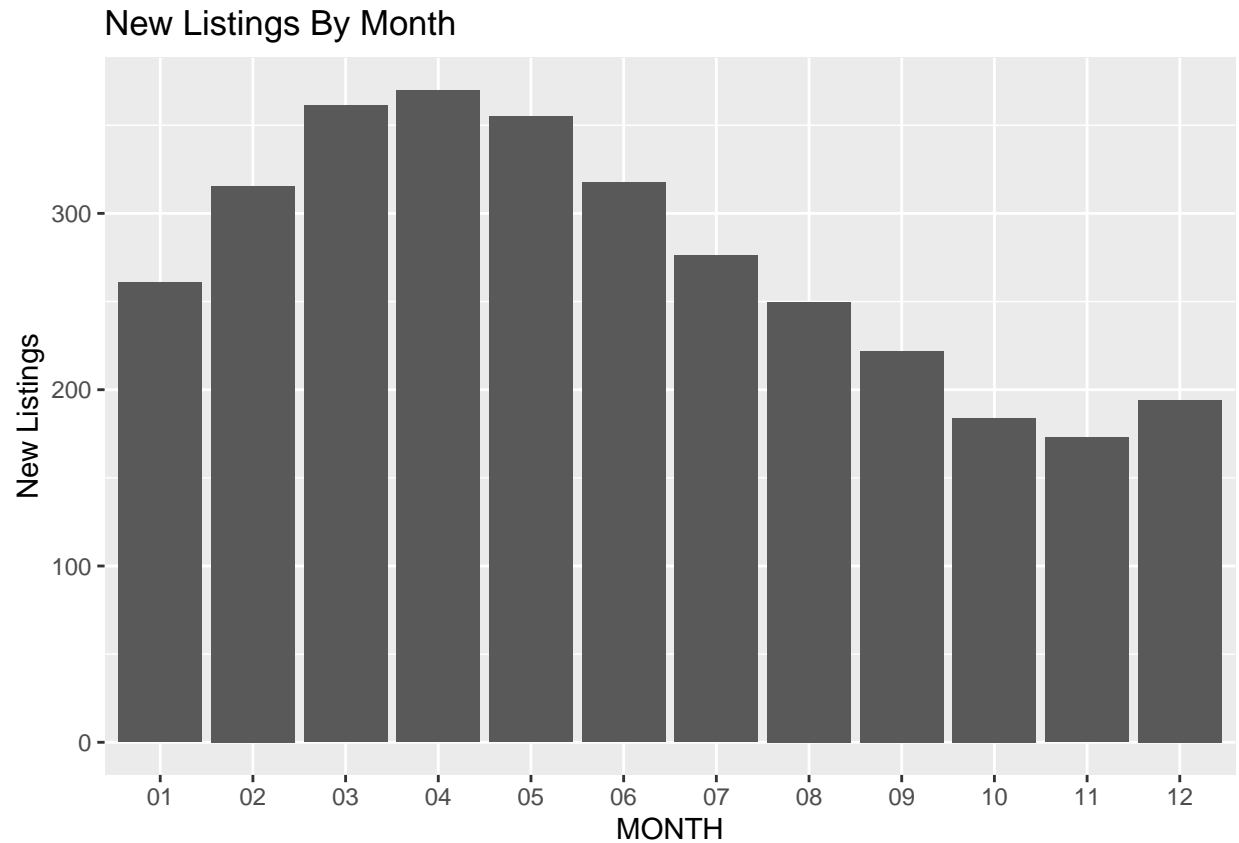


```
## # A tibble: 12 x 2
##   month Pending.Sales
##   <chr>         <dbl>
## 1 01           275.
## 2 02           313.
## 3 03           343.
## 4 04           347.
## 5 05           336.
## 6 06           312.
## 7 07           282.
## 8 08           263.
## 9 09           238.
## 10 10          215.
## 11 11          205.
## 12 12          223.
```

####New Listings Analysis

New Listings is at the highest in April, followed by March and May. New Listings is at the lowest in November.





```
## # A tibble: 12 x 2
##   month New.Listings
##   <chr>      <dbl>
## 1 01         261.
## 2 02         315.
## 3 03         361.
## 4 04         370.
## 5 05         355.
## 6 06         317.
## 7 07         276.
## 8 08         250.
## 9 09         222.
## 10 10         184.
## 11 11         173.
## 12 12         194
```

Interesting Insights and Implications to the Consumer (Target Audience)

Based on a Redfin article, it usually takes 50-60 days for a mortgage loan to close. This could explain why certain variables such as Median List Price, Inventory, Pending Sales, and New Listings seem to be generally at the highest in the months of March, April, and May, while Median Sale Price and Homes Sold are the highest in June followed by May and July.

The Redfin data supports the initial assumption that the best time to sell a home is in the Spring or Summer (March - August) in order to maximize the Sale Price. It would be best to list a home in the Spring or even early Spring (February - April) and hope to complete the sale by Summer (May - July).

Inventory is highest in April which could indicate home sellers taking advantage of potentially more buyers and possibly buyers willing to pay a premium for homes. There seems to be a strong relationship between Inventory levels and the Sale Price with respect to the time of year.

Unfortunately, the Redfin data does not include metrics on the amount of time closing took or how many offers were made and fell through to answer the question of whether the homeowner may also need to wait longer to sell to get top dollar. Observations seem to indicate that if a seller waits too long and enters the months where activity is slow and variables are lowest, that the seller may need to wait until the more promising months cycle again.

In comparing the months of different years, we see that the housing market has recovered very well since the housing crisis. It is great to sell in 2018 compared to 2011 as Median Sale Price and Median List Price have increased significantly over the years since then. There were high inventory levels in 2011 and 2012 possibly due to the lower demand for homes or higher foreclosures.

Limitations of Analysis

This analysis focused on the 30097-zip code for Duluth, GA, and neighboring zip codes. It would probably be good to compare each zip code for better insights or compare the area against the metrics for the entire state or maybe comparing different states. I had planned to use a mapping library to associate the metrics with location but did not have the time, so that could be an improvement. Additionally, other variables such as zip code median income, employment/unemployment, and other data could be other factors influencing home prices that could further be investigated.

References

Field, A. et. al. Discovering Statistics Using R. 2012

<https://www.unitedstateszipcodes.org/30097/>

<https://www.redfin.com/resources/how-long-does-it-take-to-buy-a-house>

<https://stackoverflow.com/questions/15415021/i-have-time-series-data-how-i-can-do-monthly-report-and-average-for-a-variable>

https://ro-che.info/articles/2017-02-22-group_by_month_r