Redfin Data Initial Exploratory Data Analysis (EDA)

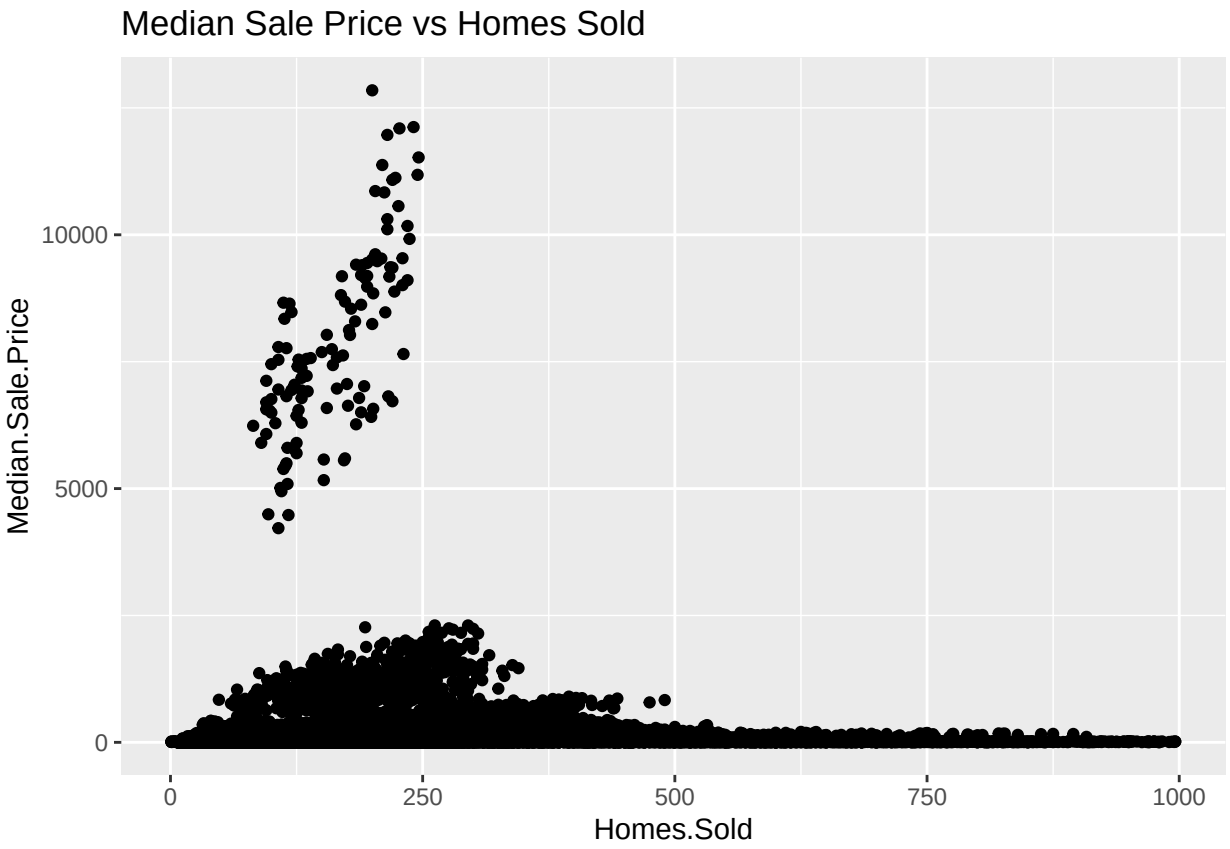
Conrad Ibanez

February 2, 2019

####a. Exploratory Graphs and Tables

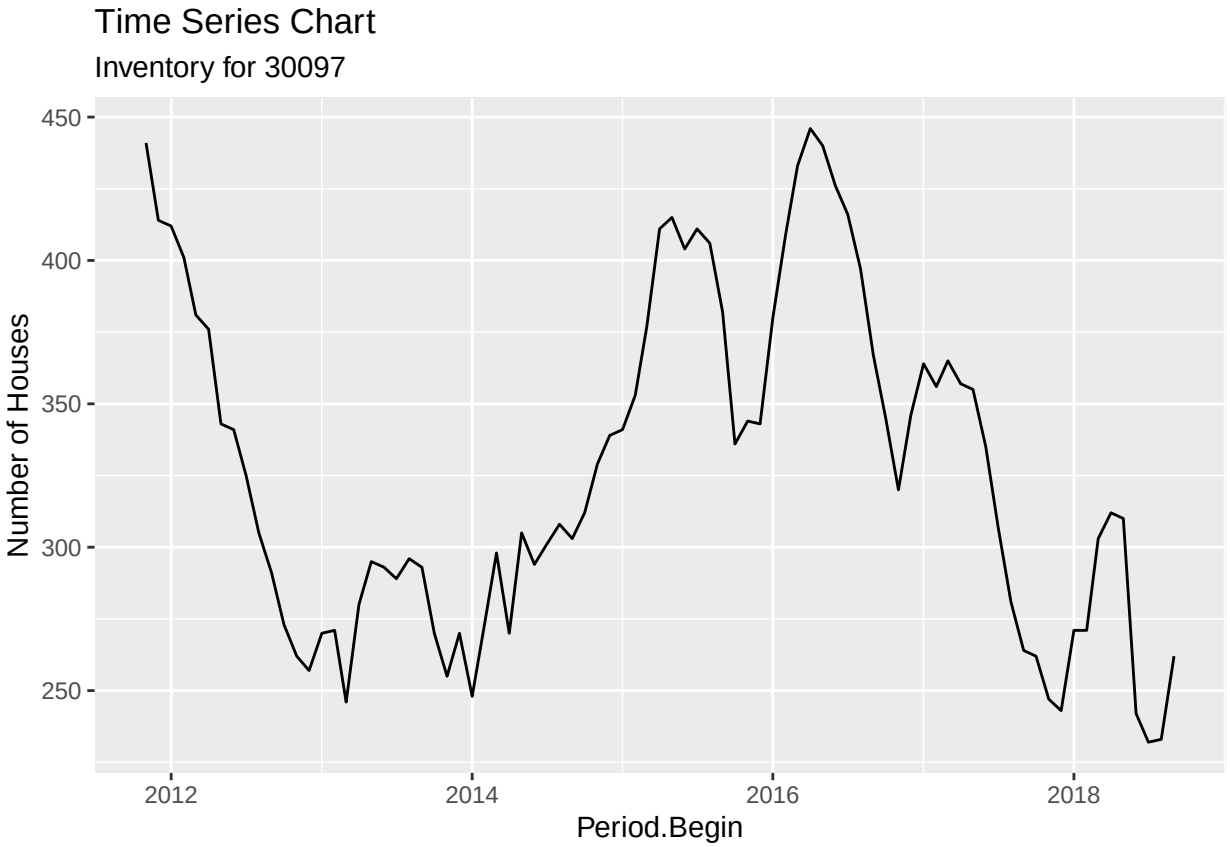
```
## 'data.frame':    48960 obs. of  55 variables:
##  $ ï..Worksheet.Filter        : chr  "Value" "Value" "Value" "Value" ...
##  $ Measure.Display            : logi  NA NA NA NA NA NA ...
##  $ Number.of.Records          : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Avg.Sale.To.List           : num  0.97 0.98 0.953 0.976 0.925 ...
##  $ Avg.Sale.To.List.Mom       : chr  "-1.0%" "-0.7%" "-1.6%" "-0.2%" ...
##  $ Avg.Sale.To.List.Yoy       : chr  "0.8%" "0.3%" "-1.0%" "0.0%" ...
##  $ City                       : chr  "" "" "" "" ...
##  $ Homes.Sold                 : int  177 99 22 278 9 206 24 213 228 9 ...
##  $ Homes.Sold.Mom             : chr  "18.8%" "-8.3%" "0.0%" "23.0%" ...
##  $ Homes.Sold.Yoy             : chr  "10.6%" "8.8%" "-31.3%" "3.0%" ...
##  $ Inventory                  : int  225 98 77 431 28 226 23 194 358 11 ...
##  $ Inventory.Mom              : chr  "-3.8%" "5.4%" "-3.8%" "2.4%" ...
##  $ Inventory.Yoy              : chr  "12.5%" "32.4%" "-14.4%" "25.7%" ...
##  $ Median.Dom                 : num  60 43 194.5 81 79.5 ...
##  $ Median.Dom.Mom             : num  -5 0 56.5 9 -43.5 9.5 11 -22 0.5 -29 ...
##  $ Median.Dom.Yoy             : num  0.5 -13 99.5 -17.5 20.5 0 -9.5 -11.5 -4.5 26 ...
##  $ Median.List.Ppsf           : num  117.5 173.8 106.5 105.6 72.1 ...
##  $ Median.List.Ppsf.Mom       : num  -0.007874 -0.000374 0.164356 -0.014055 -0.061131 ...
##  $ Median.List.Ppsf.Yoy       : num  0.1095 0.0415 0.1559 0.0382 -0.1022 ...
##  $ Median.List.Price          : num  289900 350000 222200 279900 130500 ...
##  $ Median.List.Price.Mom      : num  0 0 0.0844 -0.0509 -0.1414 ...
##  $ Median.List.Price.Yoy      : num  -0.0304 0.0495 0.0581 -0.0345 -0.1944 ...
##  $ Median.Ppsf                : num  96.7 170.1 74.4 105.7 50.8 ...
##  $ Median.Ppsf.Mom            : num  -0.042005 0.010634 -0.043392 0.000276 -0.234962 ...
##  $ Median.Ppsf.Yoy            : num  -0.0421 0.0294 -0.0784 0.0784 -0.2742 ...
##  $ Median.Sale.Price          : chr  "$197K" "$289K" "$167K" "$280K" ...
##  $ Median.Sale.Price.Mom      : chr  "-6.0%" "5.1%" "13.8%" "5.1%" ...
##  $ Median.Sale.Price.Yoy      : chr  "-24.8%" "-2.0%" "2.6%" "9.4%" ...
##  $ 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  173 83 31 355 20 145 20 268 229 10 ...
##  $ New.Listings.Mom           : chr  "-19.9%" "-2.4%" "-3.1%" "9.2%" ...
##  $ New.Listings.Yoy           : chr  "9.5%" "13.7%" "0.0%" "22.4%" ...
##  $ off_market_in_two_weeks    : num  0.1792 0.3253 0.0323 0.1577 0.1 ...
##  $ off_market_in_two_weeks_mom: num  0.02641 -0.08646 0.00101 0.01621 0.05 ...
##  $ off_market_in_two_weeks_yoy: num  -0.0297 0.065 0.0323 0.0336 0.0286 ...
##  $ pending_sales              : int  188 79 31 332 15 113 28 306 232 7 ...
```

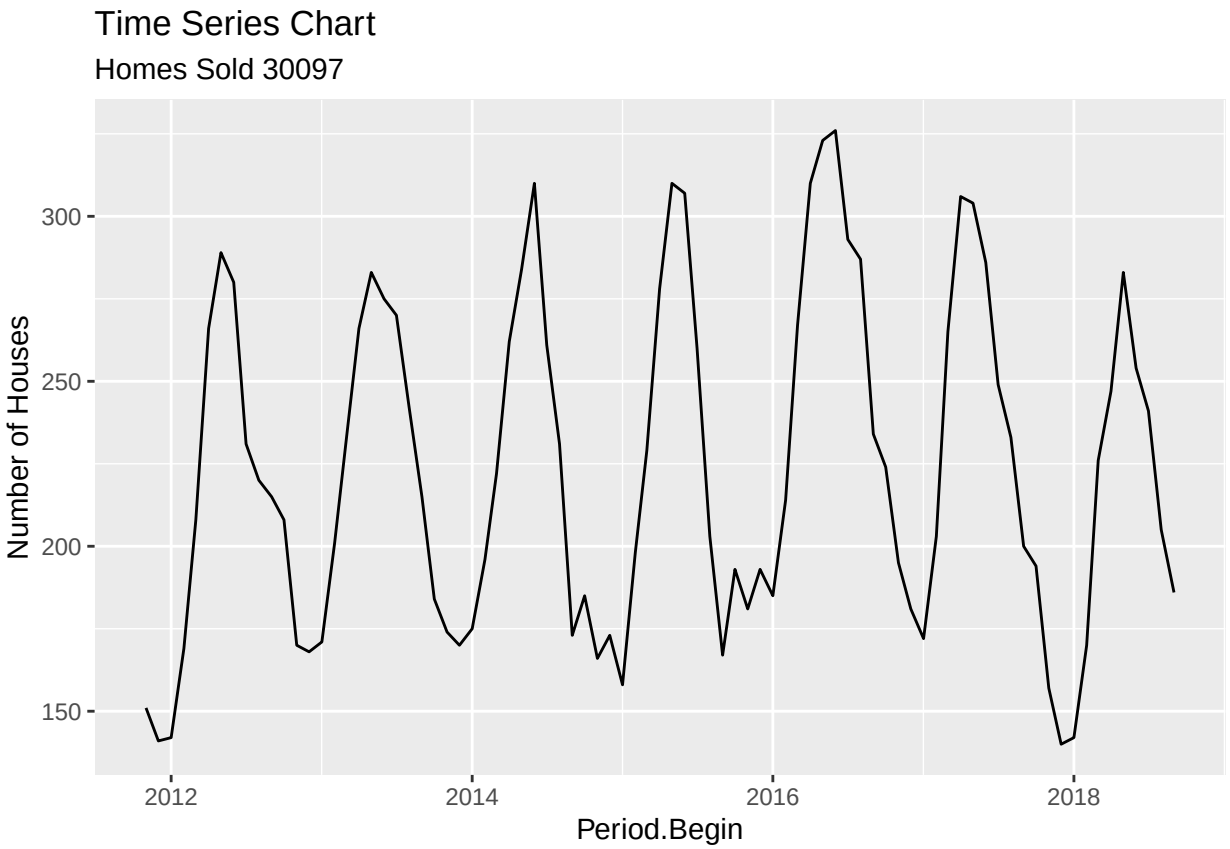
```
##  $ pending_sales_mom          : num  -0.134 -0.16 0.24 0.122 0.25 ...
##  $ pending_sales_yoy          : num  0.3239 -0.0366 0.2917 0.0153 6.5 ...
##  $ Period.Begin               : chr  "5/1/2014" "7/1/2016" "12/1/2014" "3/1/2016" ...
##  $ Period.Duration            : int  90 90 90 90 90 90 90 90 90 90 ...
##  $ Period.End                 : chr  "7/31/2014" "9/30/2016" "2/28/2015" "5/31/2016" ...
##  $ 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 Resi
##  $ Region                     : chr  "Zip Code: 30350" "Zip Code: 30329" "Zip Code: 30183" "Zip Code
##  $ Region.Type                : chr  "zip code" "zip code" "zip code" "zip code" ...
##  $ Sold.Above.List            : num  0.096 0.131 0.136 0.165 0.111 ...
##  $ Sold.Above.List.Mom        : num  -0.0449 -0.0724 -0.0455 -0.0115 0 ...
##  $ Sold.Above.List.Yoy        : num  -0.079 -0.0445 0.0114 0.0432 0.1111 ...
##  $ State                      : chr  "Georgia" "Georgia" "Georgia" "Georgia" ...
##  $ State.Code                 : chr  "GA" "GA" "GA" "GA" ...
##  $ Table.Id                   : int  12652 12632 12533 12476 12570 12389 12550 12509 12622 12535 ...

## [1] 48960

## [1] 55

## [1] 112.9739

## [1] NA

## [1] NA
```

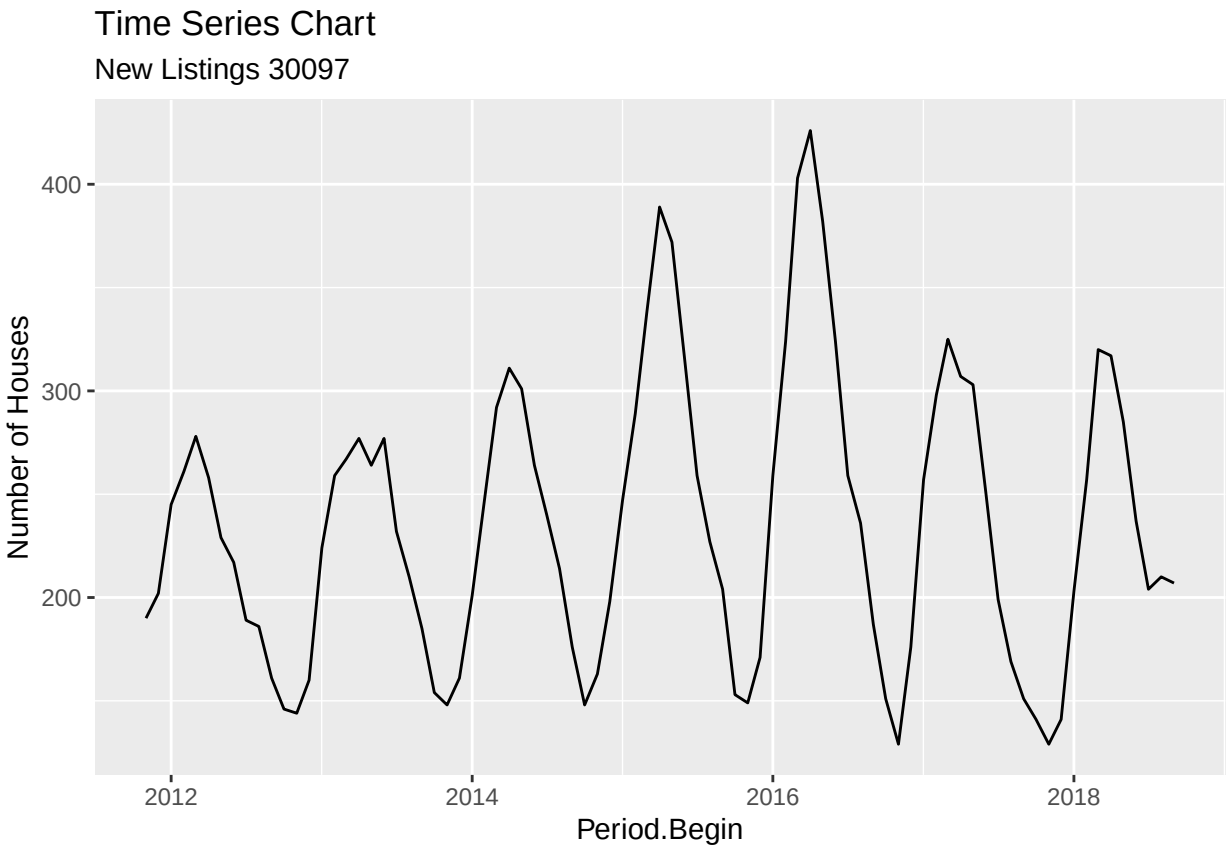

Median Sale Price vs Homes Sold
Median.Sale.Price
10000
5000
0
0
250
500
750
1000
Homes.Sold

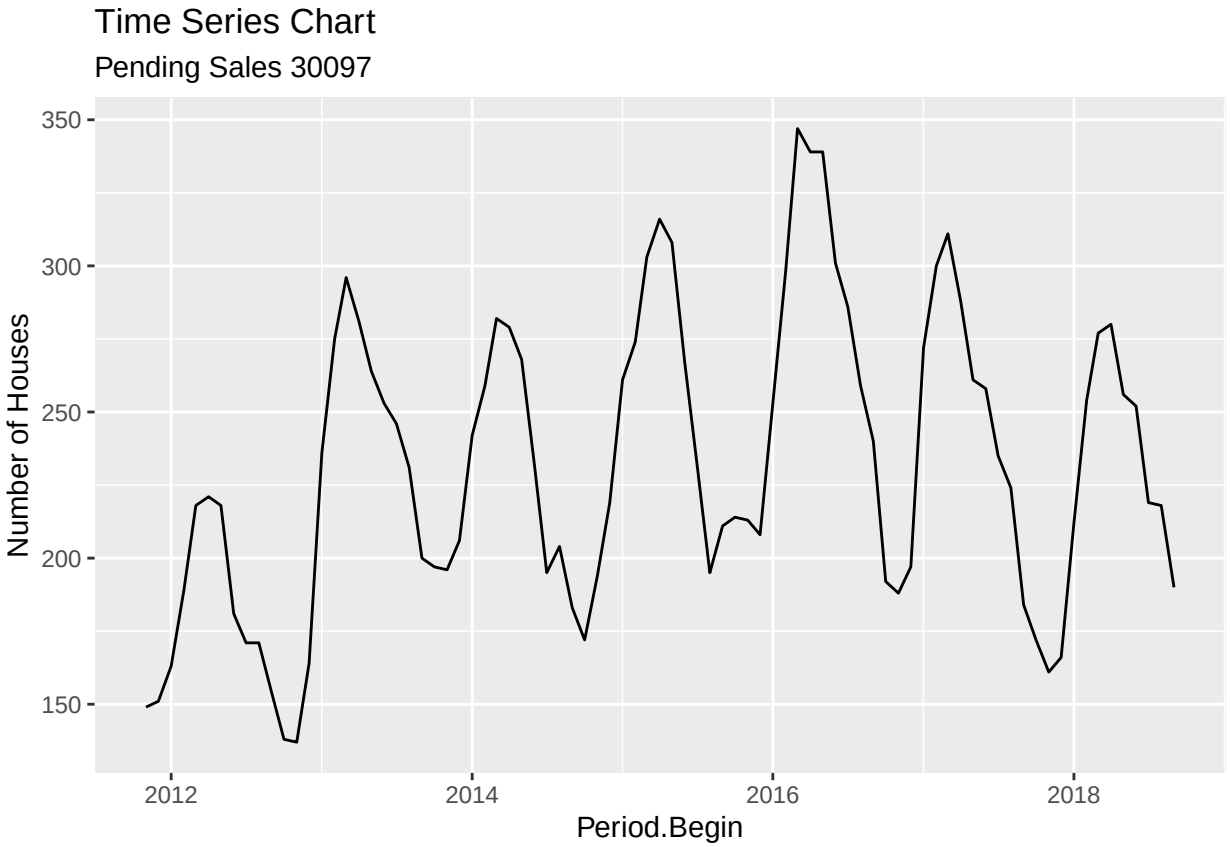

Median Sale Price vs Median List Price
Median.Sale.Price
3e+06
2e+06
1e+06
0e+00
0
250
500
750
1000
Median.List.Price

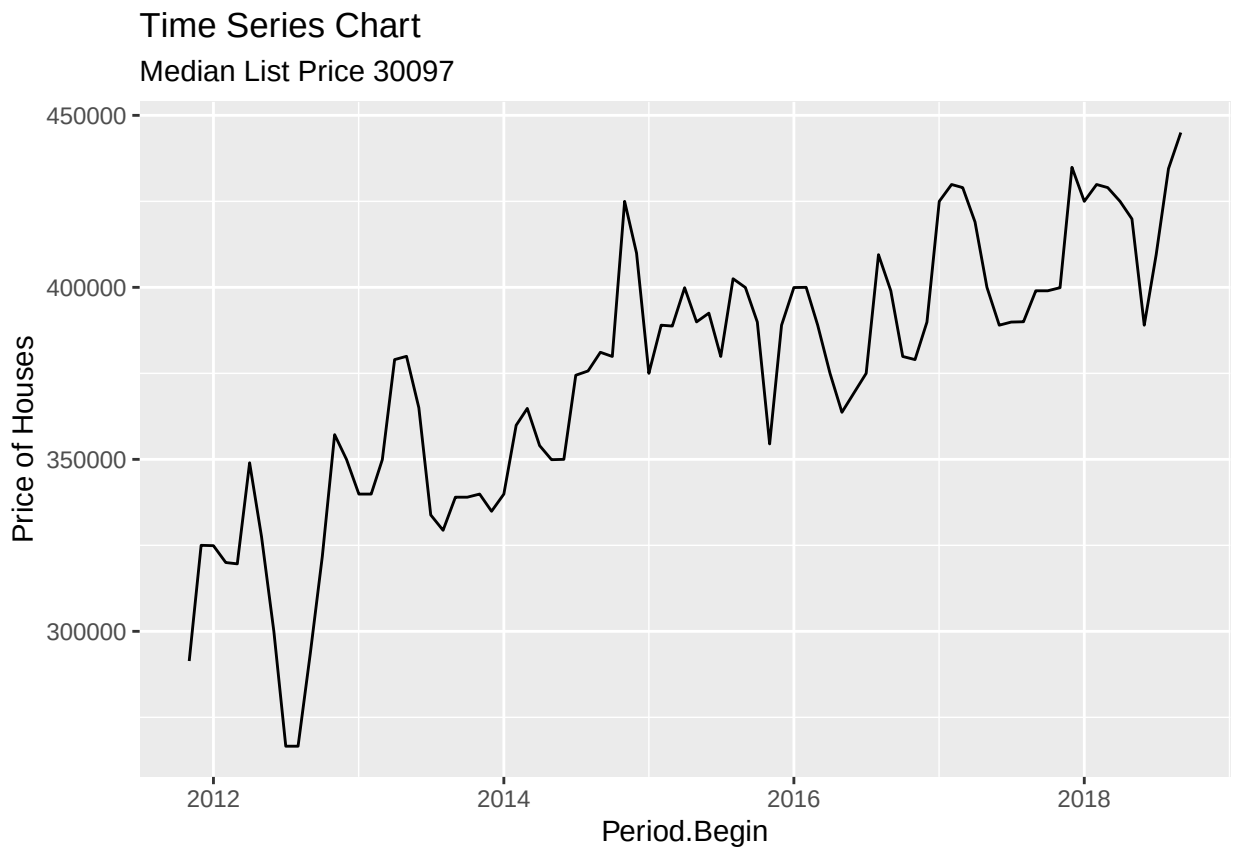
```
## [1] "data.frame"
```

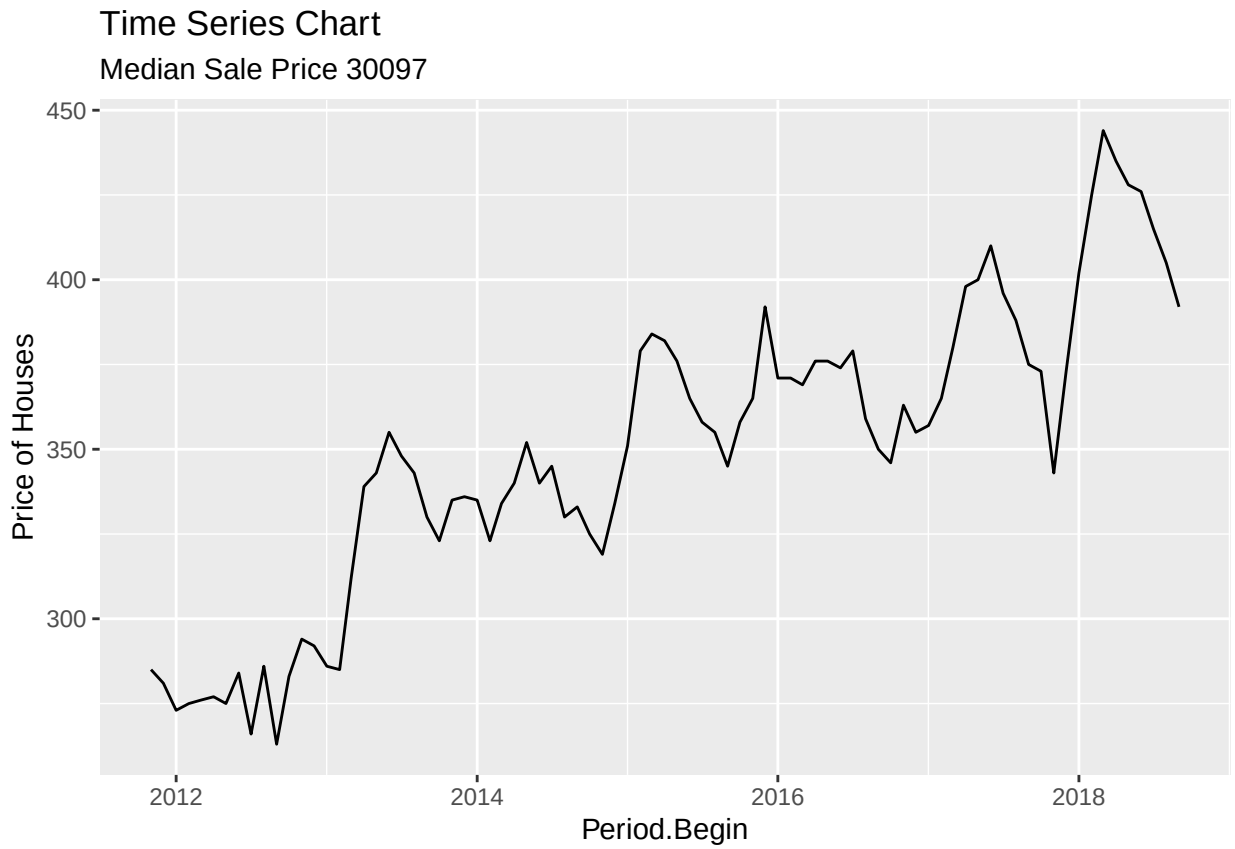

Time Series Chart
Inventory for 30097
Number of Houses
450
400
350
300
250
2012
2014
2016
2018
Period.Begin


Time Series Chart
Homes Sold 30097
Number of Houses
300
250
200
150
2012
2014
2016
2018
Period.Begin


Time Series Chart
New Listings 30097
Number of Houses
400
300
200
2012
2014
2016
2018
Period.Begin


Time Series Chart
Pending Sales 30097
Number of Houses
350
300
250
200
150
2012
2014
2016
2018
Period.Begin


Time Series Chart
Median List Price 30097
Price of Houses
450000
400000
350000
300000
2012
2014
2016
2018
Period.Begin



References

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