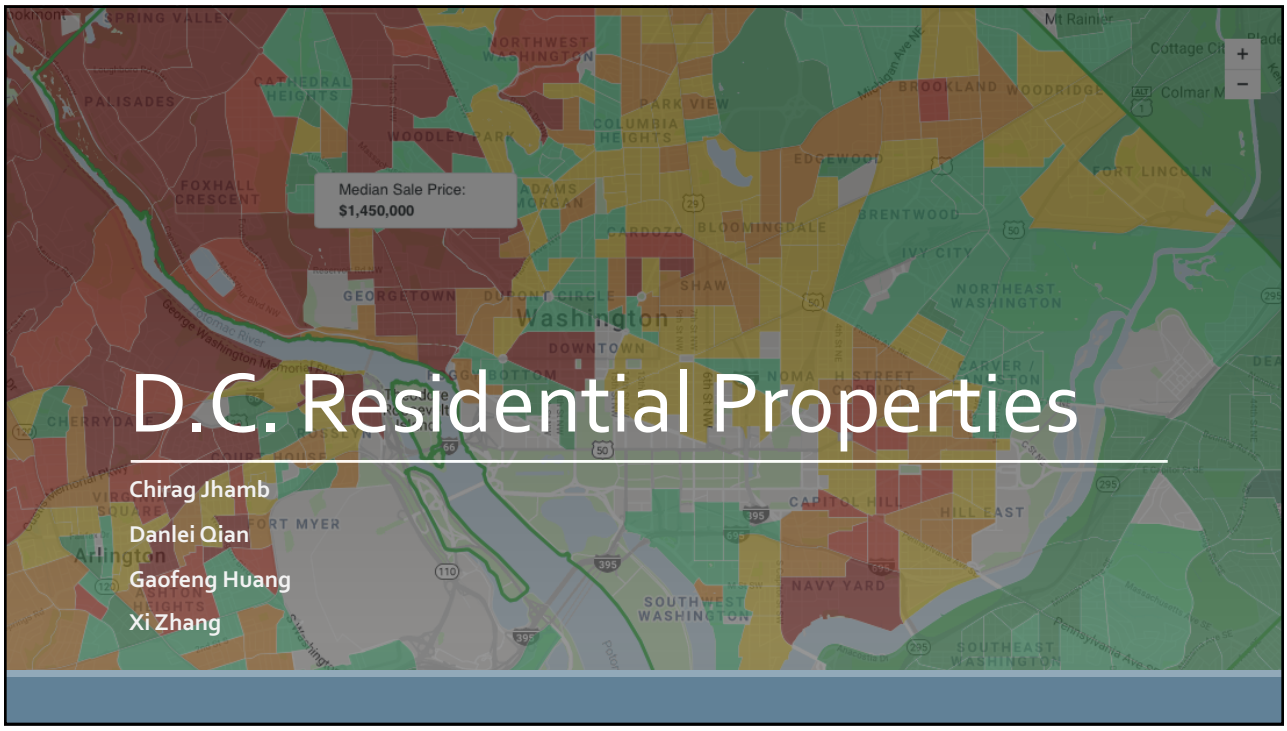


D.C. Residential Properties

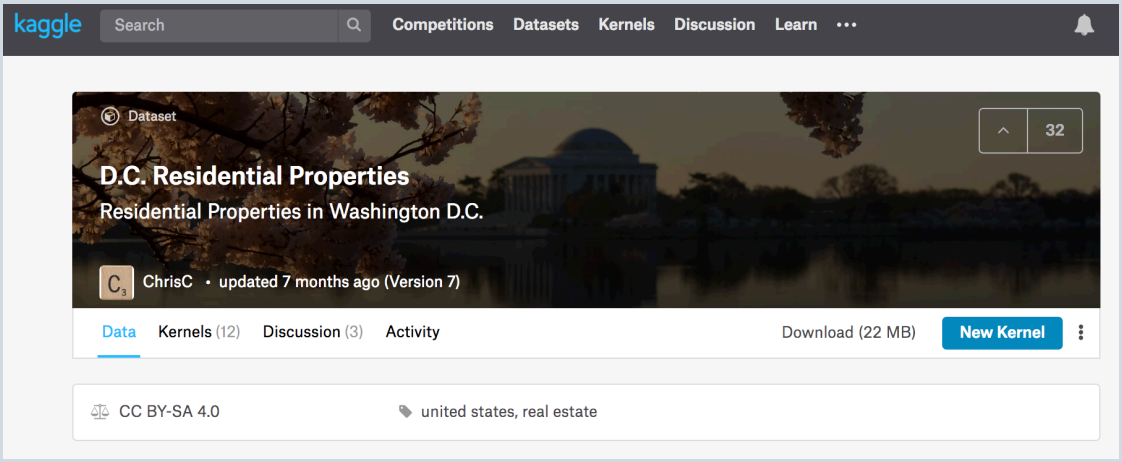
Chirag Jhamb
Danlei Qian
Gaofeng Huang
Xi Zhang

Outline

1. Description of Data
   - Data Preprocessing + Cleaning
2. Summary of Statistics
3. SMART Questions
   - Floorplan
   - Location
   - Time
   - Facility

   ] Price
4. Conclusion

Data Source



Description of RawData

```
[1] "X.1" "BATHRM" "HF_BATHRM" "HEAT"
[5] "AC" "NUM_UNITS" "ROOMS" "BEDRM"
[9] "AYB" "YR_RMDL" "EYB" "STORIES"
[13] "SALEDATE" "PRICE" "QUALIFIED" "SALE_NUM"
[17] "GBA" "BLDG_NUM" "STYLE" "STRUCT"
[21] "GRADE" "CNDTN" "EXTWALL" "ROOF"
[25] "INTWALL" "KITCHENS" "FIREPLACES" "USECODE"
[29] "LANDAREA" "GIS_LAST_MOD_DTTM" "SOURCE" "CMPLX_NUM"
[33] "LIVING_GBA" "FULLADDRESS" "CITY" "STATE"
[37] "ZIPCODE" "NATIONALGRID" "LATITUDE" "LONGITUDE"
[41] "ASSESSMENT_NBHD" "ASSESSMENT_SUBNBHD" "CENSUS_TRACT" "CENSUS_BLOCK"
[45] "WARD" "SQUARE" "X" "Y"          [49] "QUADRANT"
```

- Attributes
- Time
- Assessment
- Location

Data Preprocessing

Time:

- Turn "SALEDATE" into YYYY-format as "SALE_YR"
- "SALE_YR" - "AYB" or "YR_RMDL" or "EYB"
  → "SPAN_AYB" or "SPAN_YR_RMDL" or "SPAN_EYB"

Attribute:

- "BATHRM" = "BATHRM" + 0.5*"HF_BATHRM" (For simplicity)

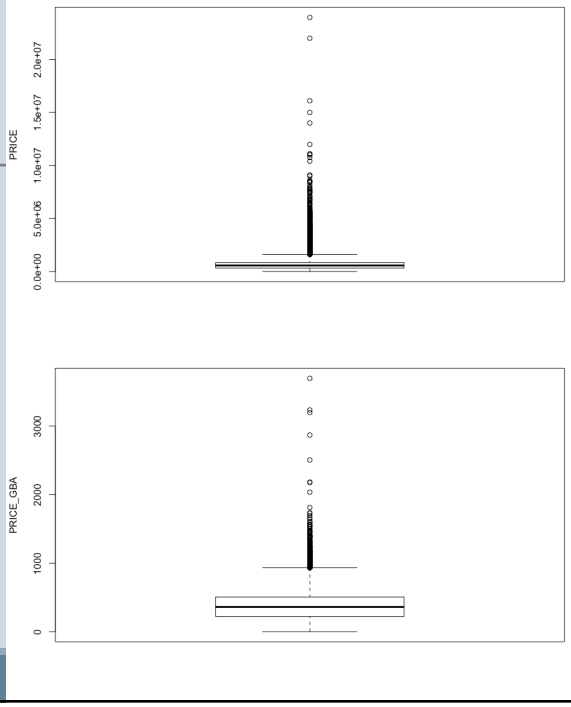
Measure:

- Set a new useful measurement
  "PRICE_GBA" = "PRICE" / "GBA" (USD per square foot)

Data Cleaning

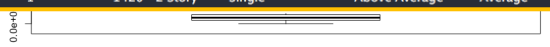
Omit NA:

- Observation: 158957 → 33162

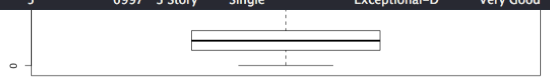


	BATH... <dbl>	HEAT <fctr>	AC <fctr>	NUM_UNITS <dbl>	ROO... <int>	BED... <int>	STORIES <dbl>	PRICE <dbl>	QUALIFIED <fctr>	SALE_NUM <fctr>	GBA <dbl>	STYLE <fctr>	STRUCT <fctr>	GRADE <fctr>	CNDTN <fctr>
81...	4.0	Hot Water Rad	Y	2	16	4	2.00	198340	U	1	3374	2 Story	Multi	Average	Average
81...	2.0	Forced Air	N	1	9	4	2.00	136000	U	1	1768	2 Story	Single	Above Average	Average
81...	1.0	Hot Water Rad	N	1	4	2	2.00	38220	U	1	1004	2 Story	Row End	Average	Average
81...	1.0	Water Base Brd	N	1	4	2	2.00	26665	Q	1	792	2 Story	Row Inside	Average	Good
81...	1.0	Forced Air	Y	1	4	2	2.00	68000	U	1	792	2 Story	Row Inside	Average	Average
81...	1.0	Hot Water Rad	N	1	6	3	2.00	93000	Q	1	1276	2 Story	Row Inside	Average	Average
81...	1.0	Hot Water Rad	N	1	5	2	2.00	53550	U	1	780	2 Story	Row Inside	Average	Average
81...	1.5	Forced Air	N	1	6	3	2.00	90000	Q	1	1088	2 Story	Row Inside	Average	Average
81...	2.0	Warm Cool	Y	1	5	3	2.00	98000	Q	1	1580	2 Story	Row Inside	Average	Average
81...	4.0	Forced Air	Y	4	16	4	2.00	150000	U	1	2720	2 Story	Multi	Average	Good
81...	4.0	Hot Water Rad	Y	4	16	4	2.00	180000	U	1	3400	2 Story	Multi	Average	Average
81...	1.5	Hot Water Rad	N	1	7	3	2.00	1	U	1	1420	2 Story	Single	Above Average	Average

Select Reasonable Data:



	BATH... <dbl>	HEAT <fctr>	AC <fctr>	NUM_UNITS <dbl>	ROO... <int>	BED... <int>	STORIES <dbl>	PRICE <dbl>	QUALIFIED <fctr>	SALE_NUM <fctr>	GBA <dbl>	STYLE <fctr>	STRUCT <fctr>	GRADE <fctr>	CNDTN <fctr>
33...	3.5	Warm Cool	Y	1	10	4	2.00	245000	Q	1	2878	2 Story	Single	Very Good	Good
33...	2.5	Forced Air	Y	1	10	4	2.00	190000	U	1	2208	2 Story	Single	Good Quality	Good
34...	2.5	Warm Cool	Y	1	11	4	2.00	244000	Q	1	3022	1.5 Stor...	Single	Good Quality	Good
35...	3.5	Hot Water Rad	Y	1	10	4	2.00	171000	Q	1	2752	2 Story	Single	Good Quality	Average
35...	3.5	Warm Cool	Y	1	10	6	2.50	285000	Q	1	3525	2.5 Stor...	Single	Very Good	Good
37...	5.5	Warm Cool	Y	1	13	6	3.00	140000...	U	4	7725	3 Story	Single	Exceptional-D	Very Good
37...	3.5	Warm Cool	Y	1	10	4	1.50	150000	U	1	3968	Split Level	Single	Above Average	Average
37...	7.5	Warm Cool	Y	1	15	6	3.00	9000000	Q	2	6164	3 Story	Single	Exceptional-B	Excellent
37...	4.0	Forced Air	Y	1	13	6	1.00	200000	U	1	2632	1 Story	Single	Good Quality	Good
38...	3.0	Forced Air	Y	1	8	4	2.50	200000	U	1	3140	2.5 Stor...	Single	Very Good	Very Good
38...	8.5	Warm Cool	Y	1	13	7	2.75	107500...	Q	5	6997	3 Story	Single	Exceptional-D	Very Good



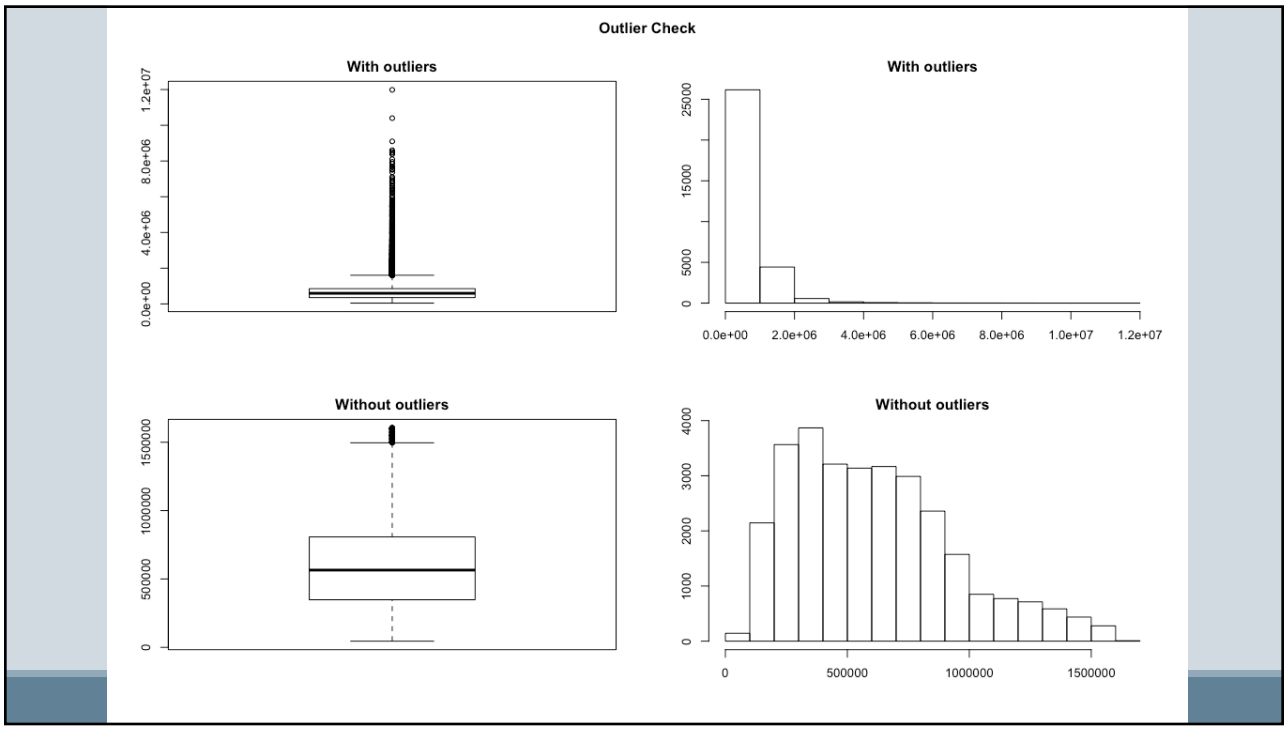
Data Cleaning

Omit NA:

- **Observation: 158957 → 33162**

Select Reasonable Data:

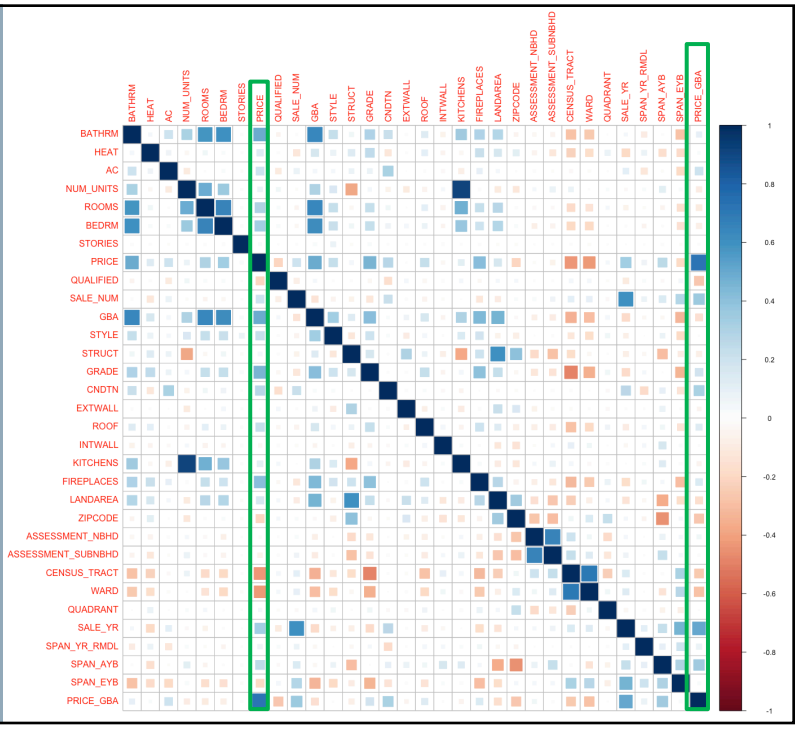
- **80 <= PRICE_GBA <= 1500 (5% - 99.9%)**
- **3 <= ROOMS <= 20 (0.1% - 99.9%)**
- **Observation: 33162 → 31435**
- **Then, use OutlierKD function w.r.t "PRICE".**
- **Observation: 31435 → 29802 (maybe omit too much large value)**



Summary of Statistics

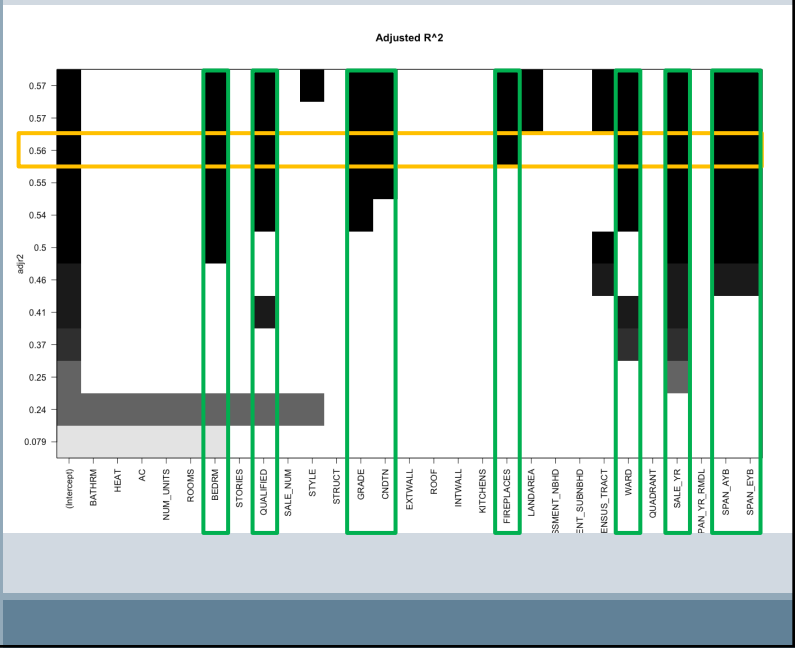
	BATHRM <dbl>	NUM_UNI... <dbl>	ROOMS <dbl>	BEDRM <dbl>	STORIES <dbl>	PRICE <dbl>	GBA <dbl>	KITCHENS <dbl>	FIREPLACES <dbl>
nbr.val	29802.000	29802.000	29802.00	29802.000	29802.00	29802.0	29802.0	29802.000	29802.000
nbr.null	1.000	23.000	0.00	4.000	6.00	0.0	0.0	14.000	14986.000
nbr.na	0.000	0.000	0.00	0.000	0.00	0.0	0.0	0.000	0.000
min	0.000	0.000	3.00	0.000	0.00	45000.0	407.0	0.000	0.000
max	11.000	4.000	20.00	15.000	826.00	1605000.0	10520.0	4.000	11.000
range	11.000	4.000	17.00	15.000	826.00	1560000.0	10113.0	4.000	11.000
sum	78162.000	36196.000	219346.00	102361.000	63209.75	18219463326.0	49736819.0	37362.000	20574.000
median	2.500	1.000	7.00	3.000	2.00	565000.0	1518.0	1.000	0.000
mean	2.623	1.215	7.36	3.435	2.12	611350.4	1668.9	1.254	0.690
SE.mean	0.006	0.003	0.01	0.006	0.03	1921.5	3.8	0.003	0.005
CI.mean.0.95	0.011	0.006	0.02	0.012	0.06	3766.3	7.5	0.007	0.010
var	1.005	0.314	4.27	1.154	25.04	110039066626.1	437503.6	0.335	0.739
std.dev	1.003	0.560	2.07	1.074	5.00	331721.4	661.4	0.579	0.860
coef.var	0.382	0.461	0.28	0.313	2.36	0.5	0.4	0.462	1.246

Correlation Map



Best Linear Model

```
PRICE_GBA ~ BEDRM + QUALIFIED +
            GRADE + CNDTN +
            FIREPLACES + WARD +
            SALE_YR + SPAN_AYB +
            SPAN_EYB
```

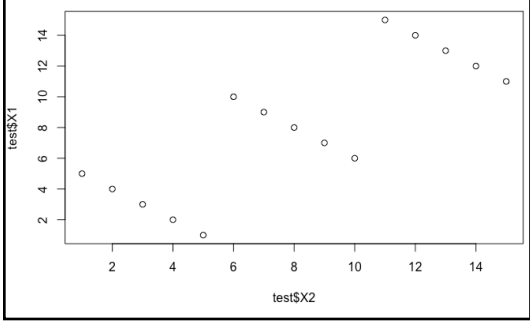


Weird? – Something about Statistics

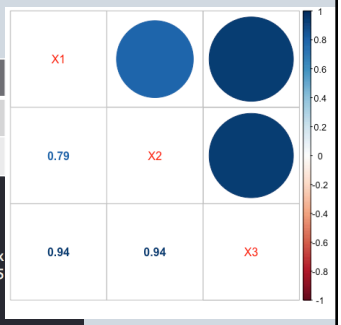
Correlation: positive <-> Coefficient (in reg.): negative

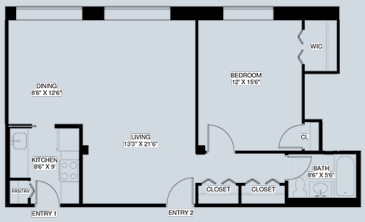
It is reasonable. An example of data below:

X1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X2	5	4	3	2	1	10	9	8	7	6	15	14	13	12	11
X3	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3

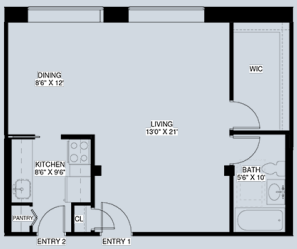



```
Call:
lm(formula = X1 ~ X2 + X3, data = test)

Residuals:
       Min         1Q     Median         3Q        Max
-3.470e-15 -1.313e-15 -2.608e-16  4.725e-16  9.652e-15

Coefficients:
             Estimate Std. Error    t value Pr(>|t|)
(Intercept) -4.000e+00  2.642e-15 -1.514e+15   <2e-16 ***
X2          -1.000e+00  6.226e-16 -1.606e+15   <2e-16 ***
X3           1.000e+01  3.295e-15  3.035e+15   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.41e-15 on 12 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 1.204e+31 on 2 and 12 DF,  p-value: < 2.2e-16
```

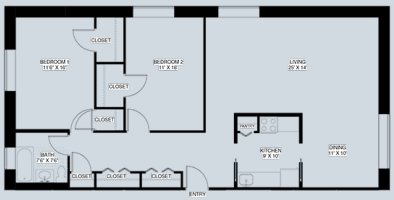




1 Bed 1 Bath



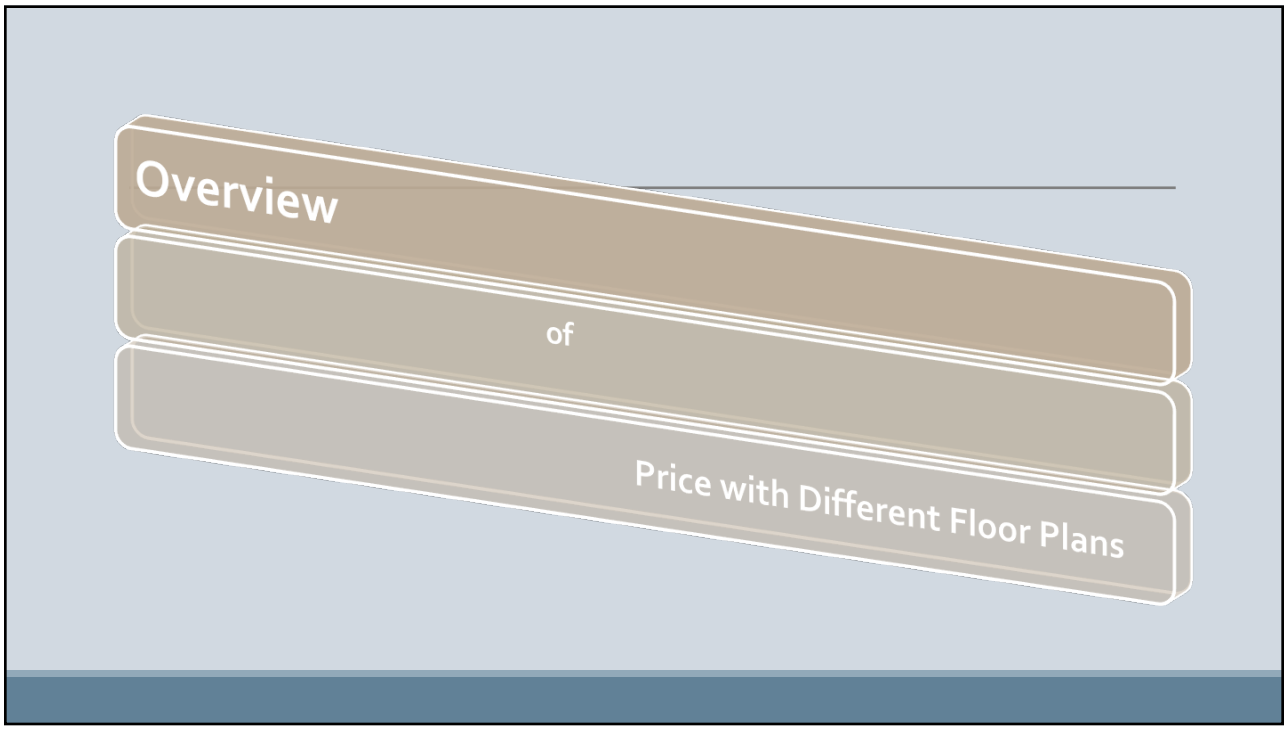
Studio



1 Bed 2 Bath

Floor Plans

How does **Floor Plan** affect **Price**?



Linear Model

Price

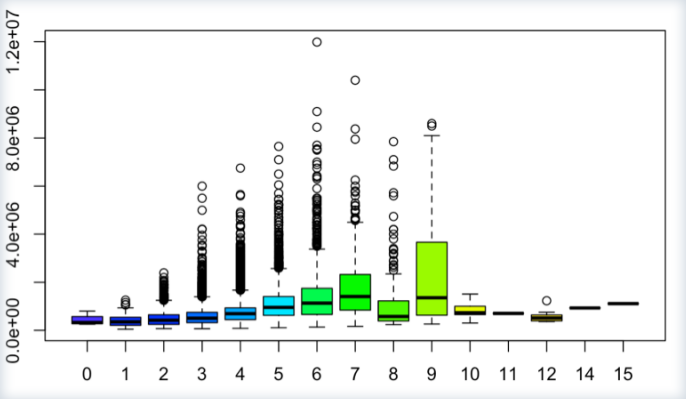
Bathroom

Bedroom

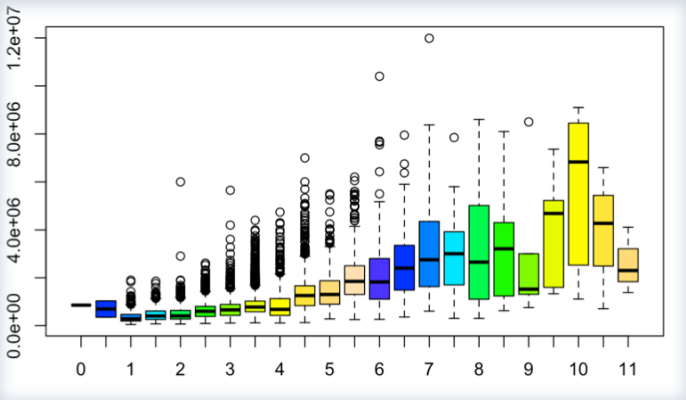
Kitchen

```
Call:
lm(formula = PRICE ~ BATHRM + BEDRM + KITCHENS, data = rp_num)

Residuals:
     Min       1Q   Median       3Q      Max
-2676540  -262056   -48967   189851 10001563

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -107405       9143  -11.75   <2e-16 ***
BATHRM        290460       3139   92.55   <2e-16 ***
BEDRM          50725       3112   16.30   <2e-16 ***
KITCHENS     -123866       4855  -25.51   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 466100 on 31431 degrees of freedom
Multiple R-squared:  0.3561,    Adjusted R-squared:  0.356
F-statistic:  5794 on 3 and 31431 DF,  p-value: < 2.2e-16

  BATHRM    BEDRM KITCHENS
1.765773 1.815708 1.152336
```

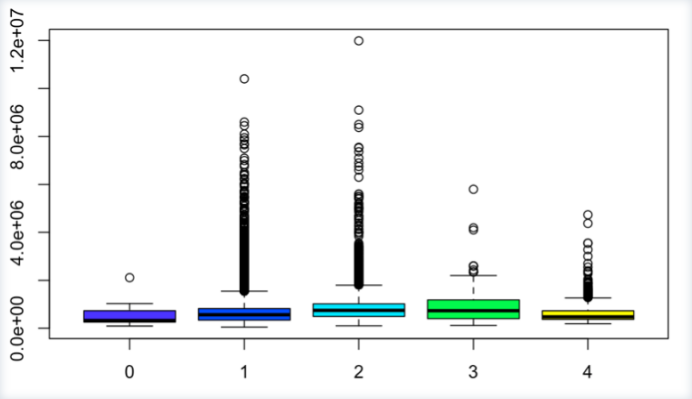
Price with different Bathrooms



Price with different Bedrooms



Price with different Kitchens

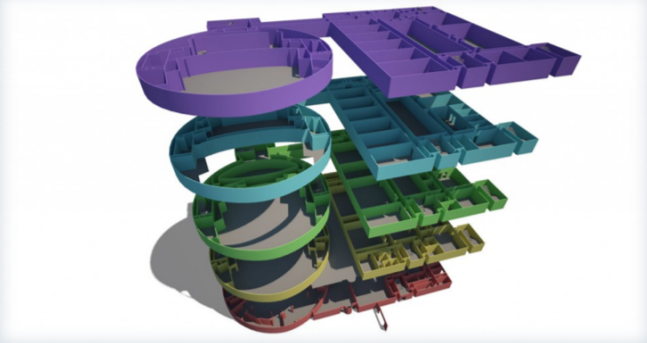


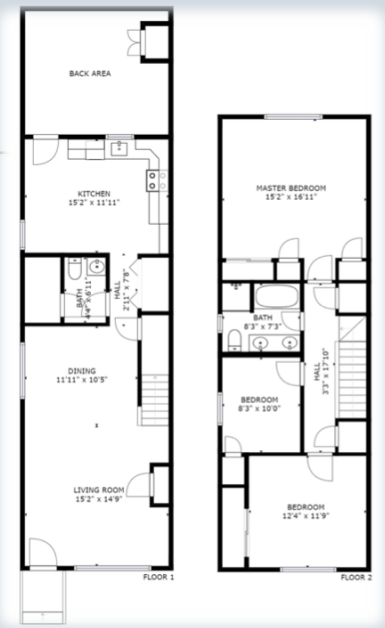
GBA and Land Area

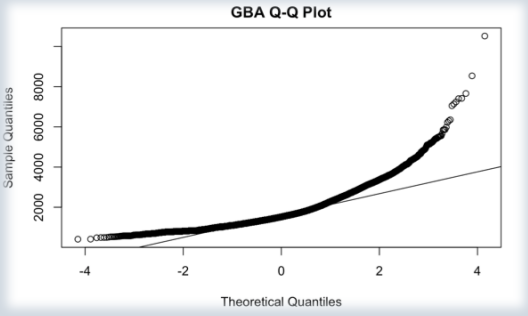
How do gba and land area influence price?

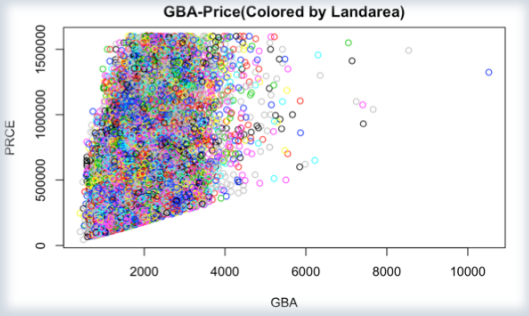
GBA: Gross Building Area.

Land Area: Land area is the area in square kilometers of the land-based portions of standard geographic areas.

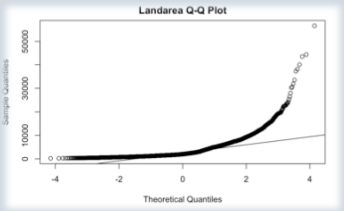


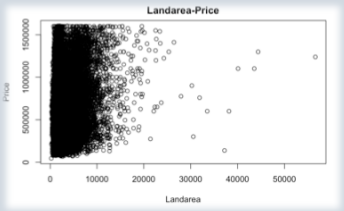


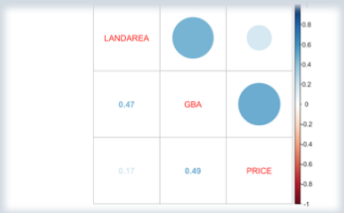




According to these two plot. Gross building area is positively related to price, and by the distribution of colors and the scatter plot of price and land area. It looks like land area doesn't affect price a lot.







These three plots may prove the weak relationship between price and land area.

ANOVA test for Land Area

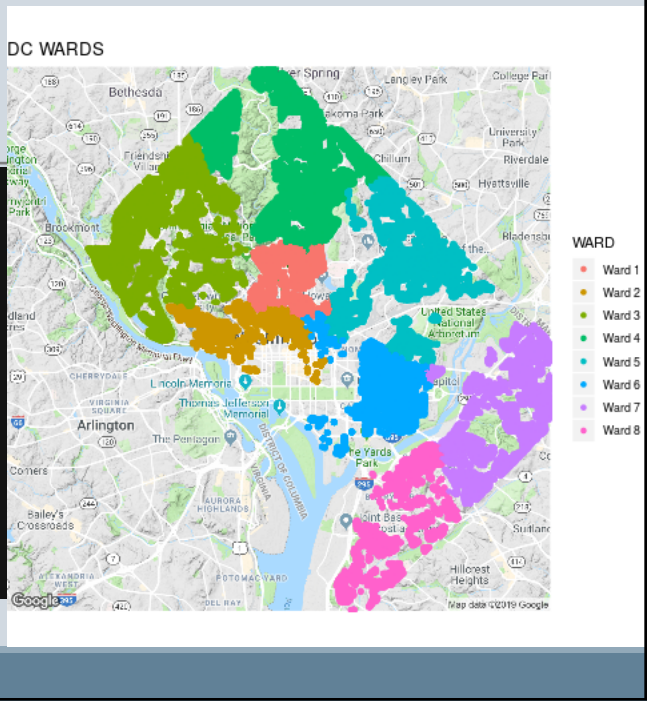
```
Analysis of Variance Table

Model 1: PRICE ~ GBA
Model 2: PRICE ~ LANDAREA + GBA
Model 3: PRICE ~ LANDAREA
  Res.Df              RSS Df         Sum of Sq        F Pr(>F)
1  31433 5667532724398895
2  31432 5667527409628765  1        5314770130     0.03   0.86
3  31433 9147628459924014 -1 -3480101050295249 19300.57 <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

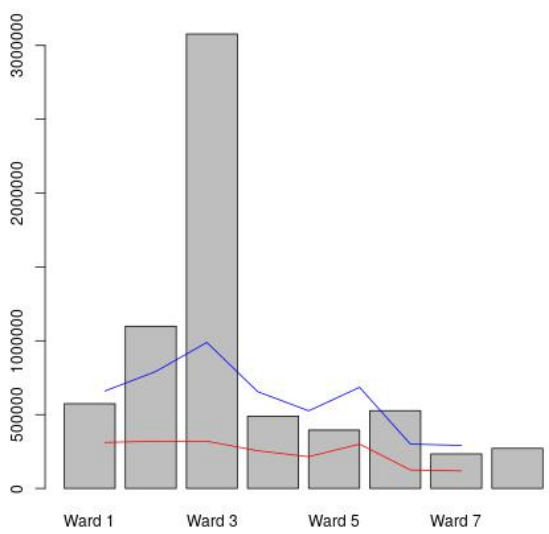
```
88    wards[[i]] <- list(sort(list(subset(data,WARD==i)$PRICE)[[1]]))
89    q1 <- quantile(wards[[i]][[1]], seq(from = 0, to = 1, by = 0.25))[[2]]
90    q2 <- quantile(wards[[i]][[1]], seq(from = 0, to = 1, by = 0.25))[[3]]
91    q3 <- quantile(wards[[i]][[1]], seq(from = 0, to = 1, by = 0.25))[[4]]
92
93    data2 <- subset(data,WARD==i)
94
95    data2$Quantile[data2$PRICE > q3] <- "4"
96    data2$Quantile[data2$PRICE < q3] <- "3"
97    data2$Quantile[data2$PRICE < q2] <- "2"
98    data2$Quantile[data2$PRICE < q1] <- "1"
99    data2 <- data2[!is.na(data2$Quantile),]
100   qmap_name <- paste(c(i ,", washington dc"), collapse = "")
101   dc <-qmap(qmap_name, zoom=15)
102   dc <- dc + geom_point(aes(x = LONGITUDE, y = LATITUDE, colour = Quantile),data = data2) + ggtitle(i)
103   cat("\nFor",i,":\n\tQuantiles are",quantile(wards[[i]][[1]], seq(from = 0, to = 1, by = 0.25)))
104   png(filename= paste(c(i,".png"), collapse = ""))
105   print(dc)
106   dev.off()
```

Location → Price

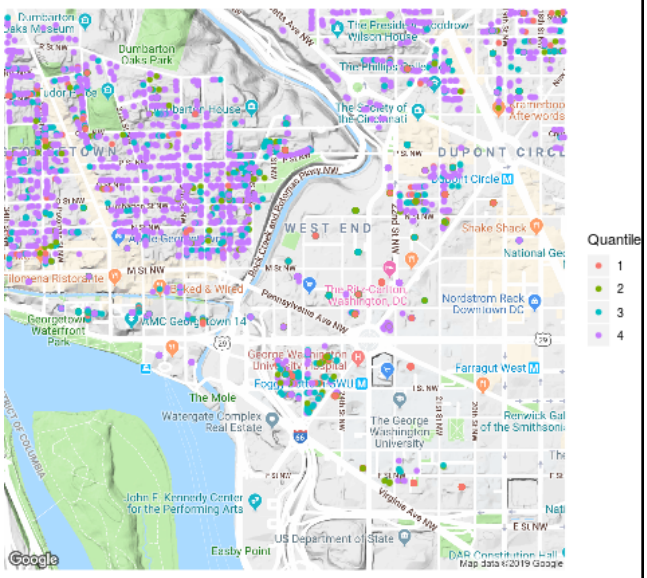
D.C. Wards and Price

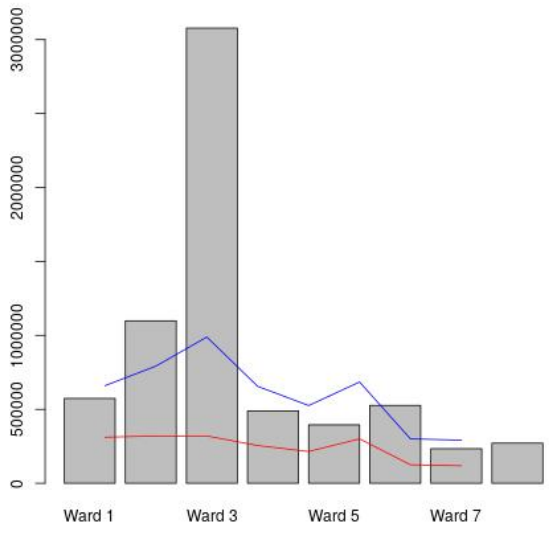
```
For Ward 1 :
        Quantiles are 10 311500 461500 660000 12780000
For Ward 2 :
        Quantiles are 10 319900 475000 792500 53969391
For Ward 3 :
        Quantiles are 1 320000 600000 989000 137427545
For Ward 4 :
        Quantiles are 250 255000 428000 655000 5000000
For Ward 5 :
        Quantiles are 1 215000 360710 525711 5250000
For Ward 6 :
        Quantiles are 1 300499 469900 685000 11000000
For Ward 7 :
        Quantiles are 10 125000 200000 300000 11000000
For Ward 8 :
        Quantiles are 1 118500 194900 292000 25100000
```



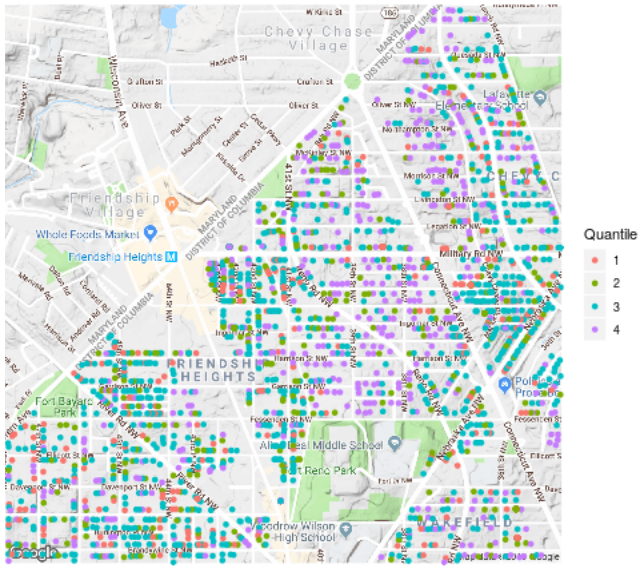

3000000
2000000
1000000
500000
0
Ward 1
Ward 3
Ward 5
Ward 7

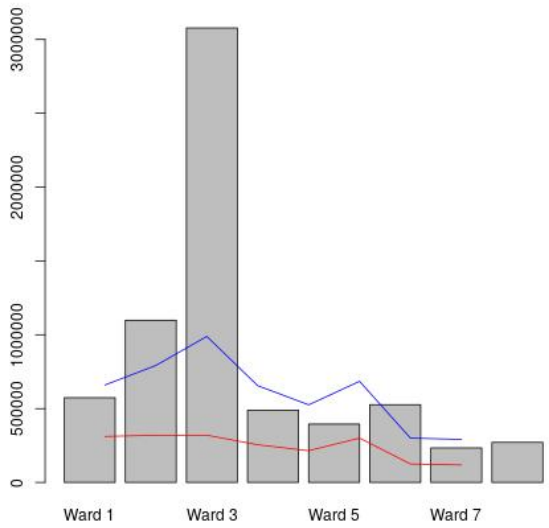
Ward 2


Dumbarton Oaks Museum
Dumbarton Oaks Park
The Phillips Collection
The Society of the Cincinnati
DUPONT CIRCLE
Dupont Circle
WEST END
Snake Shack
Nordstrom Rack Downtown DC
Farragut West
The Mole
Watergate Complex Real Estate
The George Washington University
John F. Kennedy Center for the Performing Arts
Easby Point
US Department of State
Georgetown Waterfront Park
Google
Quantile
1
2
3
4

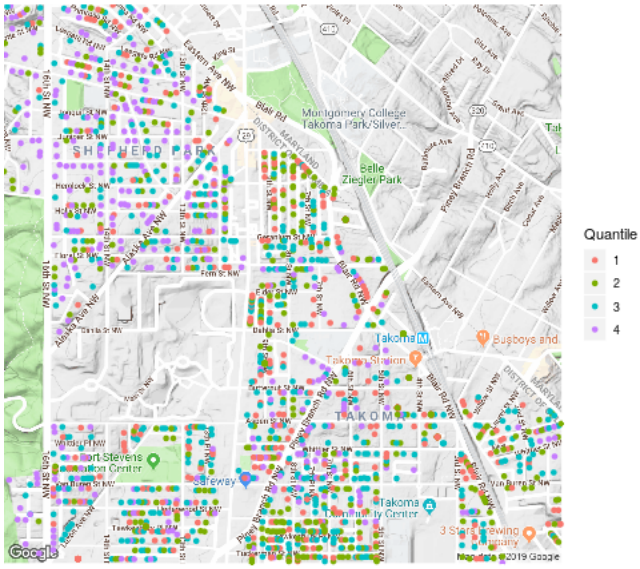

3000000
2000000
1000000
500000
0
Ward 1
Ward 3
Ward 5
Ward 7

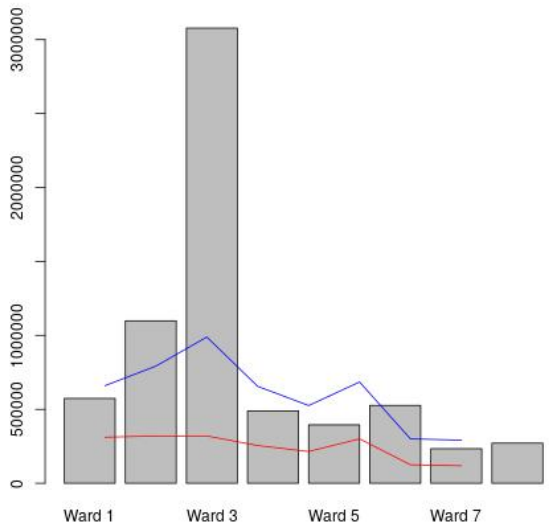
Ward 3


Chevy Chase Village
Friendship Village
Whole Foods Market
Friendship Heights
FRIENDSHIP HEIGHTS
Fort Bayard Park
Alice Deal Middle School
Fort Reno Park
Woodrow Wilson High School
Google
Quantile
1
2
3
4

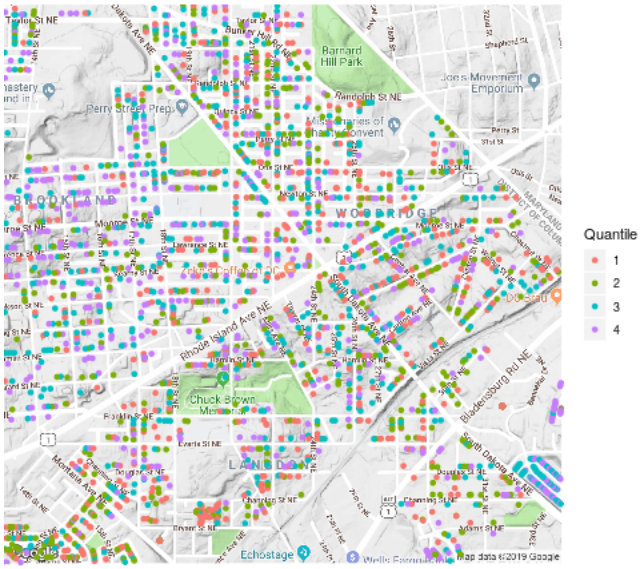

3000000
2000000
1000000
500000
0
Ward 1
Ward 3
Ward 5
Ward 7

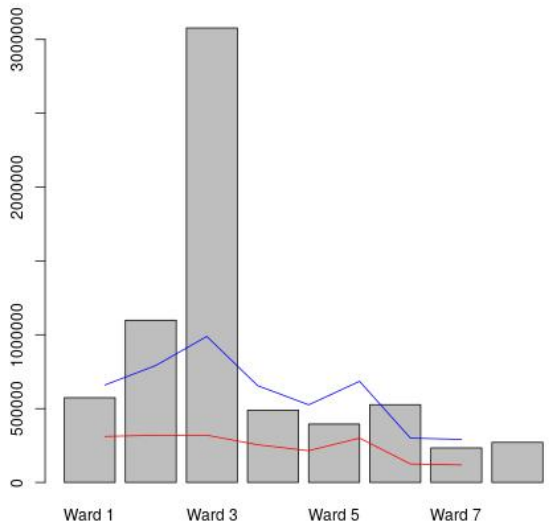
Ward 4


Quantile
1
2
3
4
Takoma
Google


3000000
2000000
1000000
500000
0
Ward 1
Ward 3
Ward 5
Ward 7

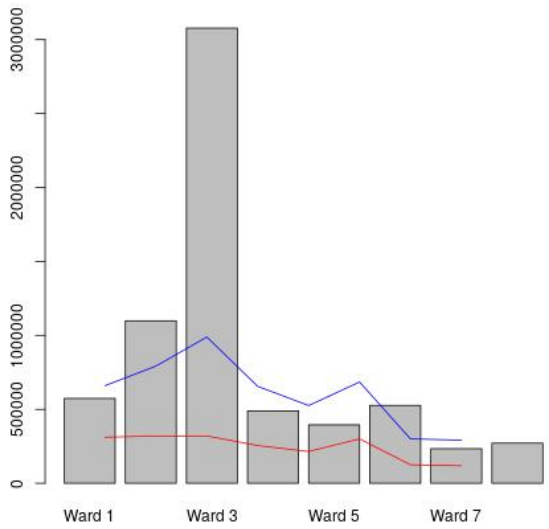
Ward 5


Quantile
1
2
3
4
Barnard Hill Park
Perry Street Prep
Chuck Brown Memorial
Echostage
Google

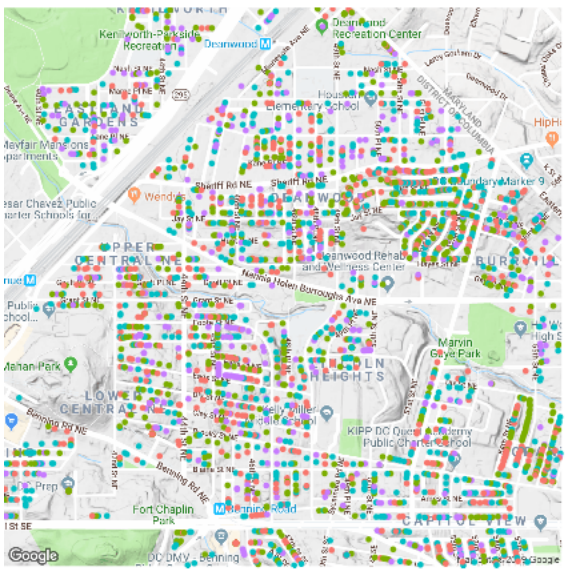

3000000
2000000
1000000
500000
0
Ward 1
Ward 3
Ward 5
Ward 7

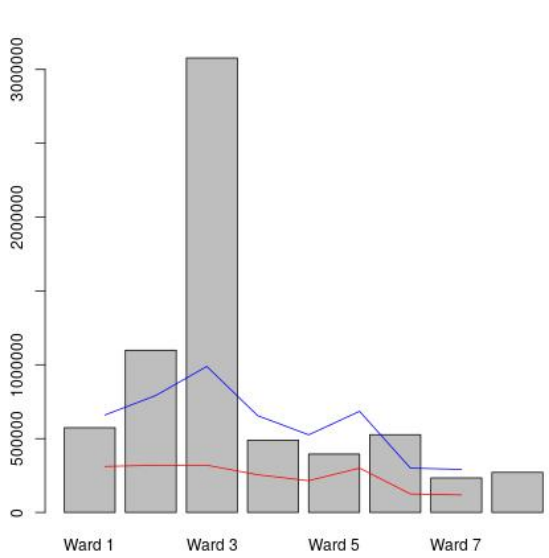
Ward 6


Botanic Garden
Independence Ave SW
Independence Ave SE
Smithsonian Institution Offices
Holiday Inn Washington-Capitol
Rayburn House Office Building
Capitol South
Federal Center SW Metro
L'Enfant Plaza
The Wharf Marina
SOUTHWEST WATERFRONT
Navy Yard-Ballpark
SOUTHWEST WASHINGTON
National's Park
Titanic Memorial
Diamond Teague Park
Fast Potomac Park
D.C. United
Washington Channel
Google
Quantile
1
2
3
4

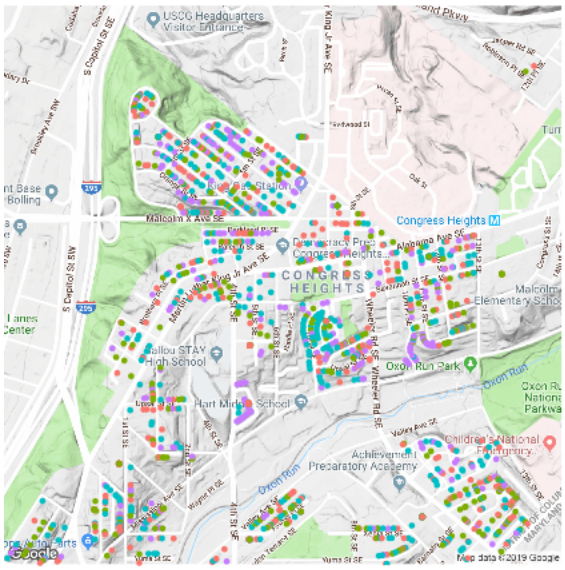

3000000
2000000
1000000
500000
0
Ward 1
Ward 3
Ward 5
Ward 7

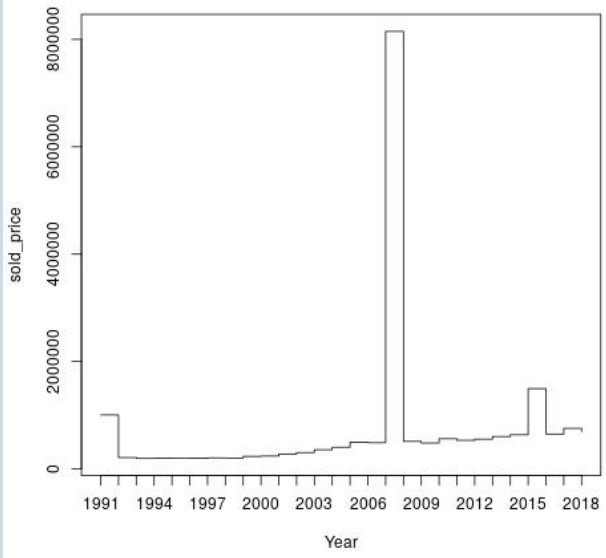
Ward 7

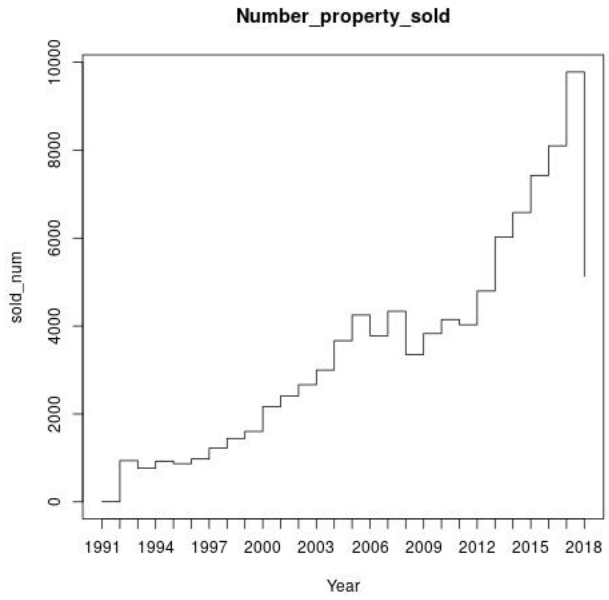

Kenilworth-Parkside Recreation
Deanwood
Deanwood Recreation Center
Elementary School
Sheriff Rd NE
Boundary Marker 9
Deanwood Rehab and Wellness Center
Marvin Gaye Park
Fort Chaplin Park
KIPP DC Quest Academy Public Charter School
Google
Quantile
1
2
3
4


3000000
2000000
1000000
500000
0
Ward 1
Ward 3
Ward 5
Ward 7

Ward 8


USCG Headquarters Visitor Entrance
Congress Heights
CONGRESS HEIGHTS
Oxon Run Park
Ballou STAY High School
Hart Middle School
Achievement Preparatory Academy
Children's National Emergency
Malcolm Elementary School
Quantile
1
2
3
4
Google


Average property price
sold_price
8000000
6000000
4000000
2000000
0
1991
1994
1997
2000
2003
2006
2009
2012
2015
2018
Year


Number_property_sold
sold_num
10000
8000
6000
4000
2000
0
1991
1994
1997
2000
2003
2006
2009
2012
2015
2018
Year

Time → Price

"SALE_NUM": Number of sale times

"YR_RMDL": Year structure was remodeled

"AYB": Year structure was built

"EYB": Year an improvement was built

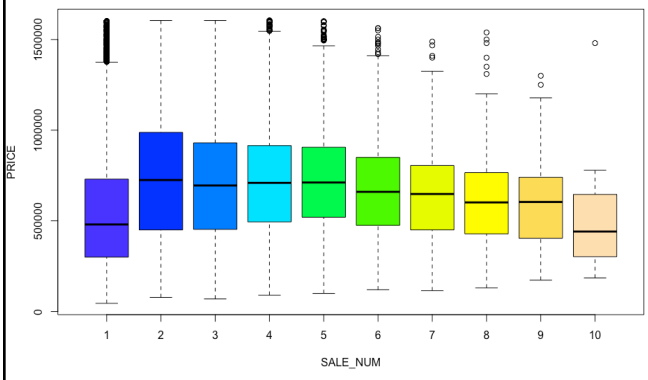
"SALE_YR": Year the property was sold

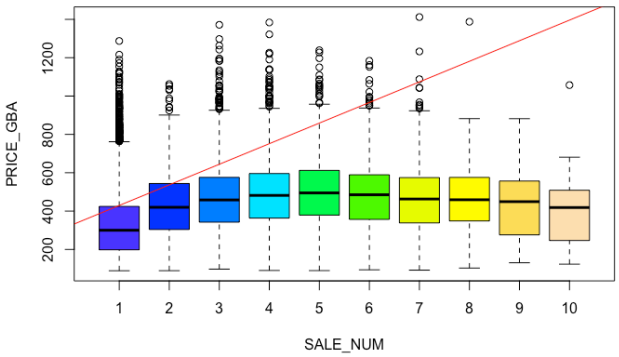
"PRICE": Price (USD)

"PRICE_GBA": Price per area (USD per sqft.)

What's the depreciation rate of price with sale number?

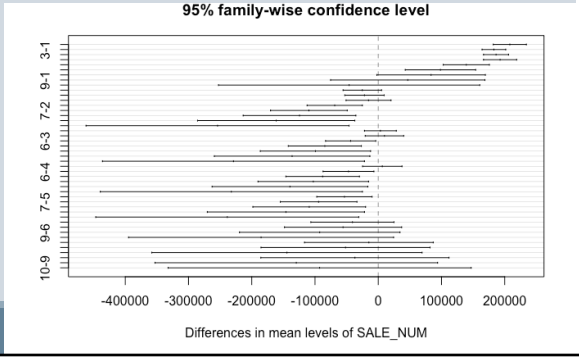
1. Boxplot
   - With regression line

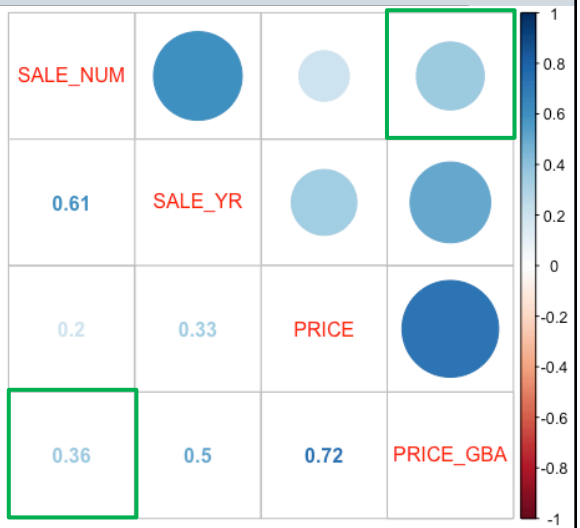




What's the depreciation rate of price with sale number?

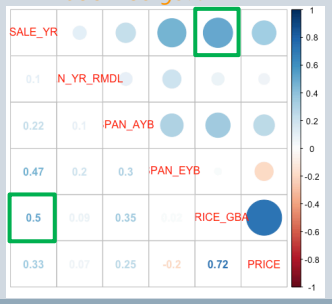
1. Boxplot
   - With regression line
2. Correlation
3. Anova test (14.4% p-values < 0.05)

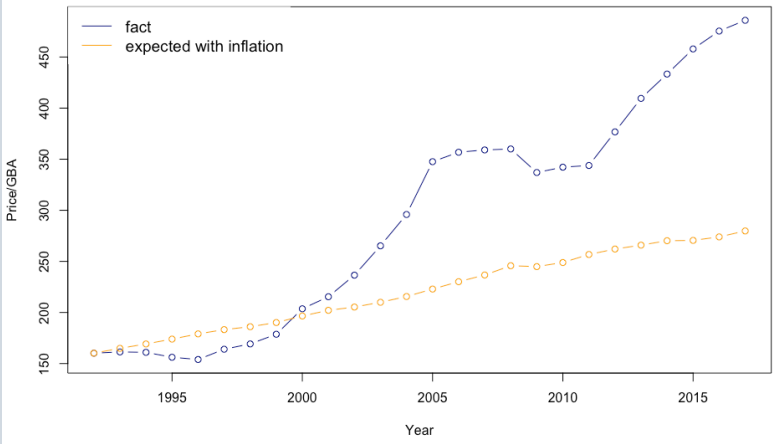




Is the price of properties in DC steady during the recent 25 years?

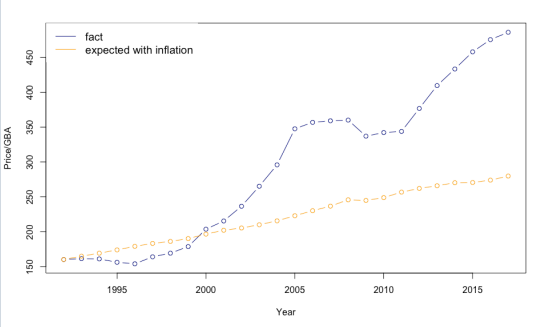
1. Correlation
   - + 0.5 moderate positive relation
2. Line chart
   - Fact: 200%+
   - Inflation: 80-90%





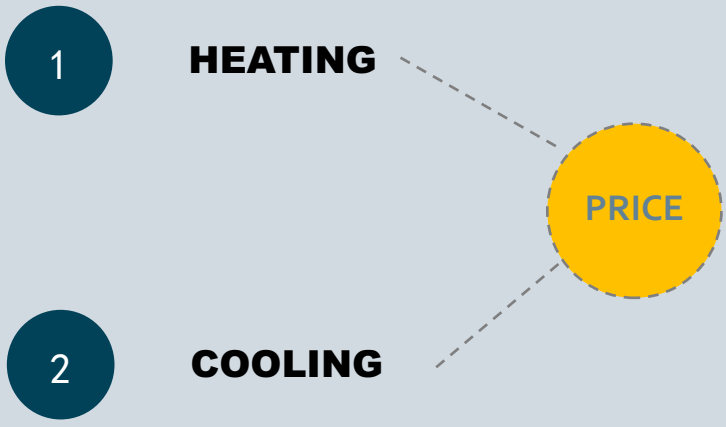
Is the price of properties in DC steady during the recent 25 years?

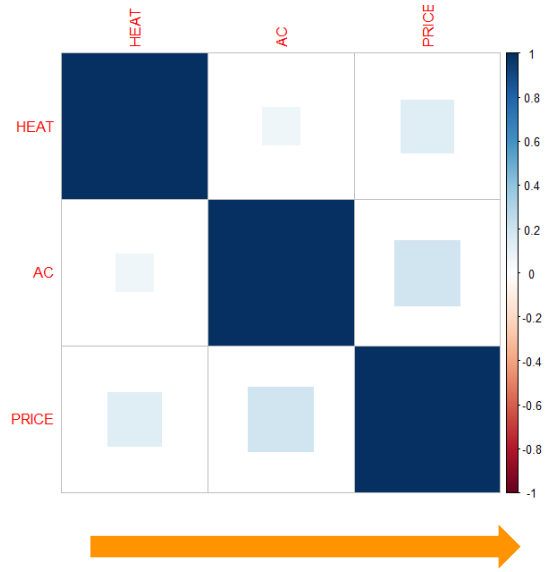
3. Two sample t-test
   - Variance test: p-value < 0.05 → diff. var.
   → Welch two sample t-test: p-value < 0.05
   → Different means




```
        F test to compare two variances

data:  yeartrend$x and fit_inf_line$value
F = 9, num df = 20, denom df = 20, p-value = 0.000001
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  3.8 19.0
sample estimates:
ratio of variances
              8.5


        Welch Two Sample t-test

data:  yeartrend$x and fit_inf_line$value
t = 3, df = 30, p-value = 0.004
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  24 119
sample estimates:
mean of x mean of y
      293       221
```

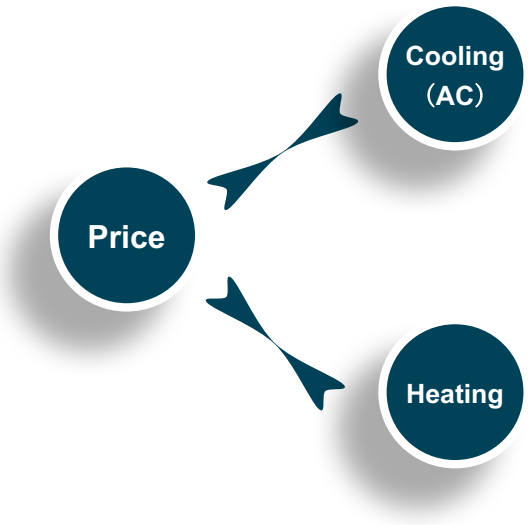
Facilities

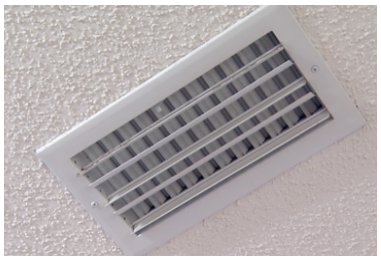




CORRELATION MATRIX

PRICE~HEAT

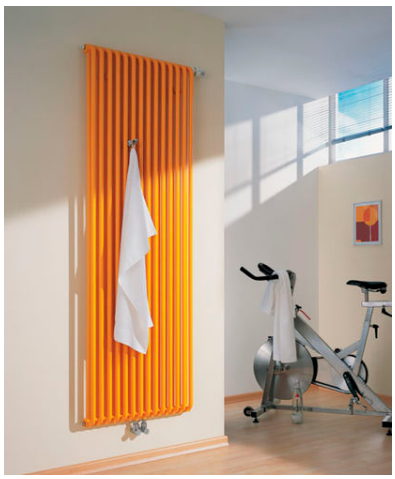




Forced Air heat

Heat Pump

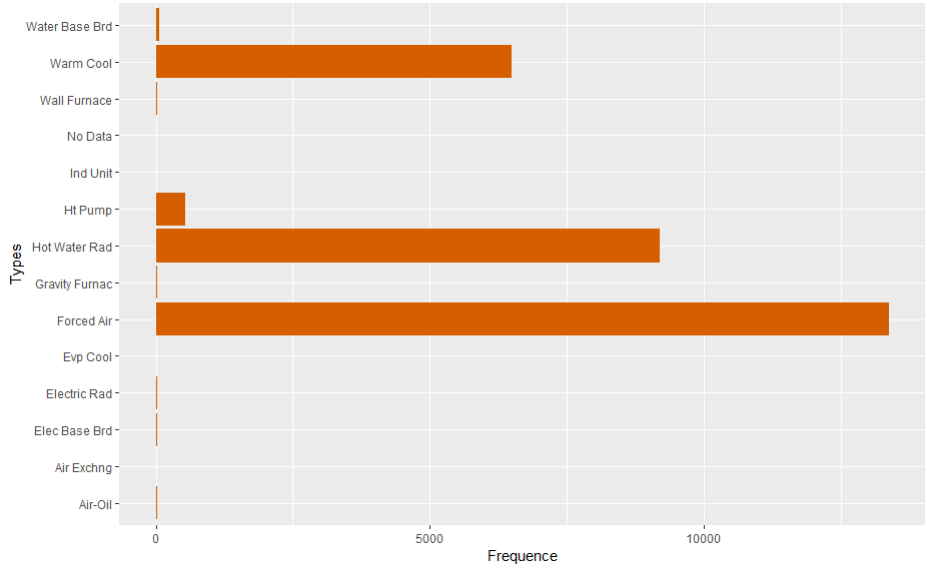
Wall Furnace



Hot Water Radiation

Heating

HEATING
BAR CHART



HEATING
ANOVA OF PRICE~HEAT

```
   aov(formula = PRICE ~ HEAT, data = HEAT)

Terms:
                       HEAT    Residuals
Sum of Squares  5.933946e+13 3.219935e+15
Deg. of Freedom           12        29789

Residual standard error: 328772.6
Estimated effects may be unbalanced
              Df    Sum Sq   Mean Sq F value Pr(>F)
HEAT          12 5.934e+13 4.945e+12   45.75 <2e-16 ***
Residuals  29789 3.220e+15 1.081e+11
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

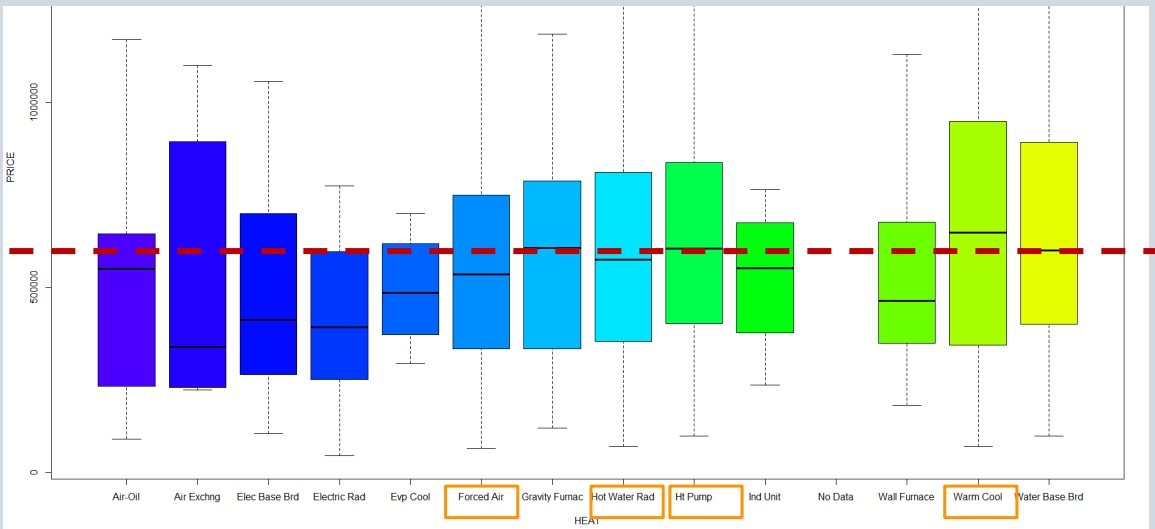
Heating facilities have significant effects on price.

?

What about the effects of different types of heaters?

HEATING

PRICE~HEAT BOXPLOT



The type of heaters **does not** have significant effect on price.

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = PRICE ~ HEAT, data = HEAT)

$HEAT
                                    diff        lwr        upr     p adj
Air Exchng-Air-Oil            -253.3333 -521638.253 521131.59 1.0000000
Elec Base Brd-Air-Oil       -24447.5000 -361611.377 312716.38 1.0000000
Electric Rad-Air-Oil       -115437.5000 -500505.782 269630.78 0.9988540
Evp Cool-Air-Oil            -26973.2143 -520530.592 466584.16 1.0000000
Forced Air-Air-Oil           51771.6509 -220675.531 324218.83 0.9999908
Gravity Furnac-Air-Oil       69181.2500 -315887.032 454249.53 0.9999951
Hot Water Rad-Air-Oil        89492.9049 -183027.950 362013.76 0.9972328
Ht Pump-Air-Oil             132501.7037 -143824.238 408827.65 0.9338754
Ind Unit-Air-Oil              4687.5000 -604158.914 613533.91 1.0000000
Wall Furnace-Air-Oil         48468.0625 -285010.852 381946.98 0.9999996
Warm Cool-Air-Oil           165718.7965 -106900.511 438338.10 0.7240572
Water Base Brd-Air-Oil      137562.7182 -171801.855 446927.29 0.9620878
Elec Base Brd-Air Exchng    -24194.1667 -511271.298 462882.96 1.0000000
Electric Rad-Air Exchng    -115184.1667 -636569.086 406200.75 0.9999518
Evp Cool-Air Exchng         -26719.8810 -632660.090 579220.33 1.0000000
Forced Air-Air Exchng        52024.9843 -392713.275 496763.24 1.0000000
Gravity Furnac-Air Exchng    69434.5833 -451950.336 590819.50 0.9999998
Hot Water Rad-Air Exchng     89746.2382 -355037.156 534529.63 0.9999821
Ht Pump-Air Exchng          132755.0371 -314369.865 579879.94 0.9989582
Ind Unit-Air Exchng           4940.8333 -698094.448 707976.11 1.0000000
Wall Furnace-Air Exchng      48721.3958 -435812.234 533255.03 1.0000000
Warm Cool-Air Exchng        165972.1298 -278871.594 610815.85 0.9911000
Water Base Brd-Air Exchng   137816.0515 -330447.860 606079.96 0.9990451
Electric Rad-Elec Base Brd  -90990.0000 -428153.877 246173.88 0.9996021
Evp Cool-Elec Base Brd       -2525.7143 -459691.748 454640.32 1.0000000
Forced Air-Elec Base Brd     76219.1509 -122852.104 275290.41 0.9888453
Gravity Furnac-Elec Base Brd 93628.7500 -243535.127 430792.63 0.9994674
Hot Water Rad-Elec Base Brd 113940.4049  -85231.665 313112.48 0.7987088
Ht Pump-Elec Base Brd       156949.2037  -47398.352 361296.76 0.3450147
Ind Unit-Elec Base Brd       29135.0000 -550602.745 608872.74 1.0000000
Wall Furnace-Elec Base Brd   72915.5625 -203869.705 349700.83 0.9996896
Warm Cool-Elec Base Brd     190166.2965   -9140.462 389473.06 0.0788694
Water Base Brd-Elec Base Brd 162010.2182  -85190.787 409211.22 0.6125482
Evp Cool-Electric Rad        88464.2857 -405093.092 582021.66 0.9999952
Forced Air-Electric Rad     167209.1509 -105238.031 439656.33 0.7109399
Gravity Furnac-Electric Rad 184618.7500 -200449.532 569687.03 0.9339419
Hot Water Rad-Electric Rad  204930.4049  -67590.450 477451.26 0.3802305
Ht Pump-Electric Rad        247939.2037  -28386.738 524265.15 0.1322186
Ind Unit-Electric Rad       120125.0000 -488721.414 728971.41 0.9999860
```

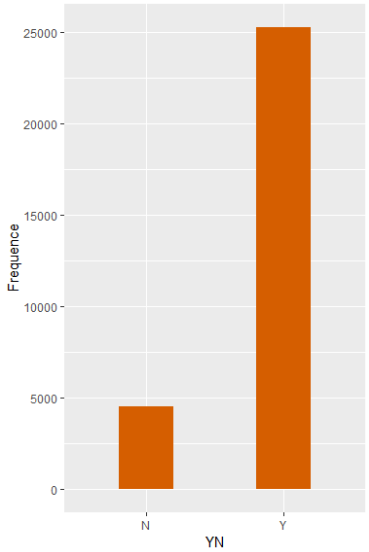
HEATING

TURKEY TEST OF PRICE~HEAT

```
Wall Furnace-Electric Rad    163905.5625 -169573.352 497384.47 0.9214489
Warm Cool-Electric Rad       281156.2965    8536.989 553775.60 0.0359312
Water Base Brd-Electric Rad  253000.2182  -56364.355 562364.79 0.2474138
Forced Air-Evp Cool           78744.8652 -333018.137 490507.87 0.9999901
Gravity Furnac-Evp Cool       96154.4643 -397402.913 589711.84 0.9999878
Hot Water Rad-Evp Cool       116466.1192 -295345.633 528277.87 0.9993593
Ht Pump-Evp Cool             159474.9180 -254864.712 573814.55 0.9883274
Ind Unit-Evp Cool             31660.7143 -650992.388 714313.82 1.0000000
Wall Furnace-Evp Cool         75441.2768 -379013.880 529896.43 0.9999980
Warm Cool-Evp Cool           192692.0107 -219184.900 604568.92 0.9445913
Water Base Brd-Evp Cool      164535.9325 -272531.273 601603.13 0.9903725
Gravity Furnac-Forced Air     17409.5991 -255037.583 289856.78 1.0000000
Hot Water Rad-Forced Air      37721.2539   22973.334  52469.17 0.0000000
Ht Pump-Forced Air            80730.0528   32710.151 128749.96 0.0000020
Ind Unit-Forced Air          -47084.1509 -591734.351 497566.05 1.0000000
Wall Furnace-Forced Air       -3303.5884 -196067.879 189460.70 1.0000000
Warm Cool-Forced Air         113947.1455   97479.856 130414.43 0.0000000
Water Base Brd-Forced Air     85791.0673  -61369.812 232951.95 0.7770919
Hot Water Rad-Gravity Furnac  20311.6549 -252209.200 292832.51 1.0000000
Ht Pump-Gravity Furnac        63320.4537 -213005.488 339646.40 0.9999285
Ind Unit-Gravity Furnac      -64493.7500 -673340.164 544352.66 1.0000000
Wall Furnace-Gravity Furnac  -20713.1875 -354192.102 312765.73 1.0000000
Warm Cool-Gravity Furnac      96537.5465 -176081.761 369156.85 0.9944253
Water Base Brd-Gravity Furnac 68381.4682 -240983.105 377746.04 0.9999515
Ht Pump-Hot Water Rad         43008.7989   -5427.343  91444.94 0.1431428
Ind Unit-Hot Water Rad       -84805.4049 -629492.461 459881.65 0.9999990
Wall Furnace-Hot Water Rad   -41024.8424 -233893.245 151843.56 0.9999680
Warm Cool-Hot Water Rad       76225.8916   58581.615  93870.17 0.0000000
Water Base Brd-Hot Water Rad  48069.8133  -99227.414 195367.04 0.9973919
Ind Unit-Ht Pump            -127814.2037 -674414.971 418786.56 0.9999112
Wall Furnace-Ht Pump         -84033.6412 -282242.194 114174.91 0.9739001
Warm Cool-Ht Pump             33217.0927  -15769.950  82204.14 0.5567901
Water Base Brd-Ht Pump         5061.0144 -149162.470 159284.50 1.0000000
Wall Furnace-Ind Unit         43780.5625 -533821.861 621382.98 1.0000000
Warm Cool-Ind Unit           161031.2965 -383705.025 705767.62 0.9990023
Water Base Brd-Ind Unit      132875.2182 -431148.551 696898.98 0.9999039
Warm Cool-Wall Furnace       117250.7340  -75756.756 310258.22 0.7249077
Water Base Brd-Wall Furnace   89094.6557 -153056.214 331245.53 0.9921341
Water Base Brd-Warm Cool     -28156.0783 -175635.378 119323.22 0.9999903
```

COOLING

PLOT



The price of properties with Air conditioner is higher than those without.

COOLING

ANOVA

```
AC = subset(rp, PRICE>0, select=c('PRICE','AC'))
aov_p_c = aov(PRICE~AC, data=AC)
aov_p_c
summary(aov_p_c)
```

```
Call:
   aov(formula = PRICE ~ AC, data = AC)

Terms:
                         AC    Residuals
Sum of Squares  1.285228e+14 3.150751e+15
Deg. of Freedom            1        29800

Residual standard error: 325161.4
Estimated effects may be unbalanced
              Df    Sum Sq   Mean Sq F value Pr(>F)
AC             1 1.285e+14 1.285e+14    1216 <2e-16 ***
Residuals  29800 3.151e+15 1.057e+11
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Cooling facilities(Air conditioners) have significant effects on price.

P-value <0.001

COOLING
T-TEST

H_0:Whether properties have or have no air conditioners , the means of the price are equal.
H_1: H_0:When properties have or have no air conditioners , the means of the price equal are **not** equal.

Construct t-intervals at 0.999 level.

```
AC_Y<-subset(rp_num,AC==2)
AC_N<-subset(rp_num,AC==1)
t.test(AC_Y$PRICE, mu = mean(AC_N$PRICE), conf.level = 0.999)
```

```
        One Sample t-test

data:  AC_Y$PRICE
t = 86.766, df = 25280, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 456059
99.9 percent confidence interval:
 632177.8 646064.4
sample estimates:
mean of x
 639121.1
```

Reject H_0 !

Cooling facilities have significant effects on price.

Conclusion

- Data cleaning is very important before we analyze data
- SMART Questions

1. Floorplan: Gross building area is a significant factor.
2. Location: Certain areas (ward 2&3) have higher price.
3. Time: Growth of properties is twice than the inflation rate
4. Facility: Both cooling and heating affect price, but types of heating do not matter to the price.

Thank you! | Q&A