

CC08E2: Group Project

Kunal Ostwal - 500668171

2022-10-05

Group Contract

Group number: CC08E2

Name: Kunal Ostwal - 500668171

GitHub link: <https://github.sydney.edu.au/kost6112/CC08E2.git>

I agree to:

- Abide by the terms of this contract in relation to the group assessment for DATA2002/2902.
- Store all my written and code contributions to the assessment in the GitHub repository.
- Keep a record of my other contributions to the assessment (e.g. discussions, emails, meetings attended). A copy of this may be requested by the coordinator.
- The group will maintain a promote high academic standars, commit to a culture of integrity, and support academic freedom.
- The group will help encourage and where possible help other members when difficulties with the work may arise.
- Work collaboratively and contribute equitably to group work, projects and other learning experiences
- Communication to be taken place over created WhatsApp group and during scheduled group meetings.
- All members will ensure notifications for WhatsApp group are on and will read and respond if required to all recieved messages.
- Daily updates provided by group members on the progress of the work allocated to them.
- Weekly meetings during during lab tutorials and/or extra informal meetings as any issues arise during the course of the project.

I understand that:

- My agreement to these terms is indicated through the act of submitting this in Canvas.
- If I fail to meet my obligations as detailed in this group contract then I have failed to meet the assessment requirements for DATA2002/2902 and may be awarded a mark of zero for some or all of the project components.

Exploratory Data Analysis

Dataset: House Prices of New York

Variable of Interest: Age

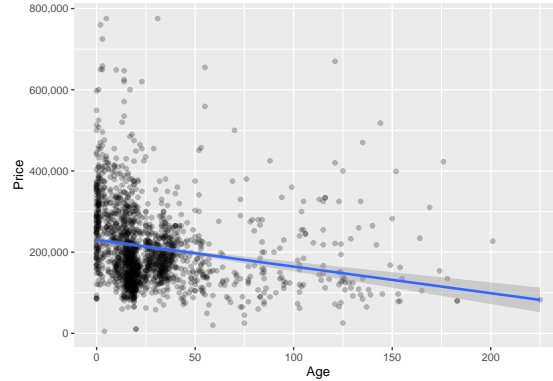
```
str(houses)
```

```
## 'data.frame':    1734 obs. of  17 variables:
## $ Price          : int  132500 181115 109000 155000 86060 120000 153000 170000 90000 122900 ...
## $ Lot.Size       : num   0.09 0.92 0.19 0.41 0.11 0.68 0.4 1.21 0.83 1.94 ...
## $ Waterfront     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ Age            : int   42 0 133 13 0 31 33 23 36 4 ...
## $ Land.Value     : int   50000 22300 7300 18700 15000 14000 23300 14600 22200 21200 ...
## $ New.Construct: Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...
## $ Central.Air    : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...
## $ Fuel.Type      : Factor w/ 7 levels "Electric","Gas",...: 1 2 2 2 2 2 4 4 1 2 ...
## $ Heat.Type      : Factor w/ 4 levels "Electric","Hot Air",...: 1 3 3 2 2 2 3 2 1 2 ...
## $ Sewer.Type     : Factor w/ 3 levels "None/Unknown",...: 2 2 3 2 3 2 2 2 1 ...
## $ Living.Area    : int   906 1953 1944 1944 840 1152 2752 1662 1632 1416 ...
## $ Pct.College    : int   35 51 51 51 51 22 51 35 51 44 ...
## $ Bedrooms       : int    2 3 4 3 2 4 4 4 3 3 ...
## $ Fireplaces     : int    1 0 1 1 0 1 1 1 0 0 ...
## $ Bathrooms      : num    1 2.5 1 1.5 1 1 1.5 1.5 1.5 1.5 ...
## $ Rooms          : int    5 6 8 5 3 8 8 9 8 6 ...
## $ Test           : Factor w/ 2 levels "0","1": 1 2 2 1 1 2 1 1 1 2 ...
```

```
describe(houses[, c("Price" , "Lot.Size", "Age", "Land.Value", "Living.Area", "Pct.College", "Bedrooms",
```

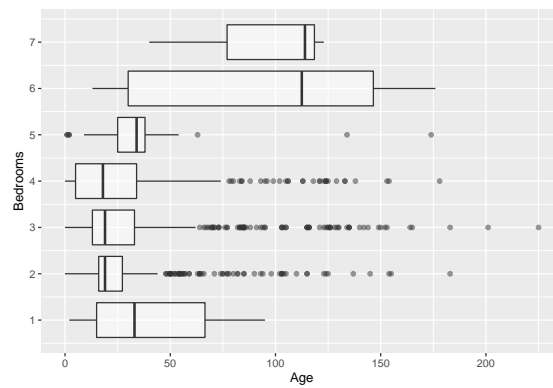
```
##          vars    n      mean      sd min      max      range      se
## Price          1 1734 211545.05 98553.81 5000 775000.0 770000.0 2366.73
## Lot.Size       2 1734      0.50    0.70   0      12.2      12.2    0.02
## Age           3 1734     28.26    29.86   0      225.0     225.0    0.72
## Land.Value     4 1734   34536.23 34980.94 200 412600.0 412400.0  840.05
## Living.Area    5 1734    1752.63  620.22 616    5228.0    4612.0   14.89
## Pct.College    6 1734     55.57   10.32  20     82.0     62.0    0.25
## Bedrooms       7 1734      3.15    0.82   1       7.0       6.0    0.02
## Fireplaces     8 1734      0.60    0.56   0       4.0       4.0    0.01
## Bathrooms      9 1734      1.90    0.66   0       4.5       4.5    0.02
## Rooms         10 1734      7.03    2.32   2      12.0      10.0    0.06
```

Age vs Price



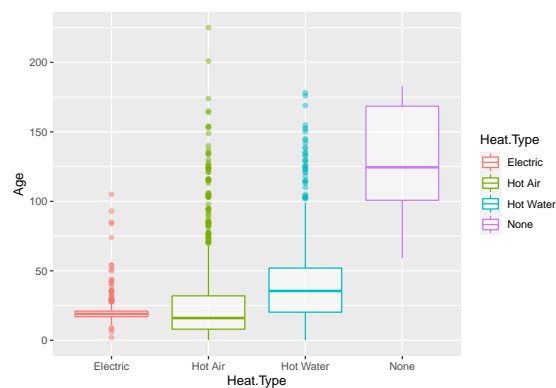
Although at the risk of extrapolation, one can infer that age and price are inversely related.

Age vs Bedrooms



We observe that the houses with 5+ bedrooms are comparatively older than <5 bedrooms houses. For less than 5 bedroom houses, the trend is opposite. Four bedroom houses have been built more recently than 1 bedroom houses.

Age vs Heat Type



We observe here that houses with no heating system are older than ones that have a heating system.

```
houses[, -which(sapply(df, class) == "factor")]
```

```
## data frame with 0 columns and 1734 rows
```

```
houses_cor_mat = houses[, -which(sapply(df, class) == "factor")] %>%  
  data.frame() %>%  
  rownames_to_column(var = "var1") %>%  
  gather(key = "var2", value = "cor", -var1)
```

Analyzing the factors that determine the price of the house

Assumptions

1. Linearity

```
M0 = lm(Price ~ 1, data = houses)
```

```
M1 = lm(Price ~ ., data = houses)
```

```
round(summary(M1)$coef, 3)
```

##	Estimate	Std. Error	t	value	Pr(> t)
## (Intercept)	7994.062	19604.816	0.408	0.684	
## Lot.Size	7554.997	2238.108	3.376	0.001	
## Waterfront1	120237.097	15524.380	7.745	0.000	
## Age	-134.154	58.115	-2.308	0.021	
## Land.Value	0.927	0.048	19.491	0.000	
## New.Construct1	-45650.453	7299.341	-6.254	0.000	
## Central.Air1	9927.179	3474.258	2.857	0.004	
## Fuel.TypeGas	10223.278	12128.326	0.843	0.399	
## Fuel.TypeNone	-30023.683	34670.469	-0.866	0.387	
## Fuel.TypeOil	6419.419	12869.599	0.499	0.618	
## Fuel.TypeSolar	-4203.574	58643.987	-0.072	0.943	
## Fuel.TypeUnknown/Other	-53945.471	58525.780	-0.922	0.357	
## Fuel.TypeWood	-49251.610	58706.645	-0.839	0.402	
## Heat.TypeHot Air	686.336	12310.746	0.056	0.956	
## Heat.TypeHot Water	-9566.646	12835.126	-0.745	0.456	
## Sewer.TypePrivate	4294.846	17107.562	0.251	0.802	
## Sewer.TypePublic	2942.735	17051.032	0.173	0.863	
## Living.Area	69.774	4.610	15.136	0.000	
## Pct.College	-107.882	151.308	-0.713	0.476	
## Bedrooms	-7659.045	2565.934	-2.985	0.003	
## Fireplaces	1010.917	2982.794	0.339	0.735	
## Bathrooms	23082.766	3361.301	6.867	0.000	
## Rooms	3012.511	960.905	3.135	0.002	
## Test1	5629.548	3245.333	1.735	0.083	

```

res = bind_rows(broom::glance(M1),
                broom::glance(M0))

res$model = c("All", "Null")
res %>% pivot_longer(
  cols = -model,
  names_to = "metric",
  values_to = "value"
) %>% pivot_wider(names_from = "model") %>%
  gt::gt() %>%
  gt::fmt_number(columns = 2:3,
                 decimals = 2) %>%
  gt::fmt_missing()

```

```

## Warning: Since gt v0.6.0 the 'fmt_missing()' function is deprecated and will soon be
## removed.
## * Use the 'sub_missing()' function instead.

```

metric	All	Null
r.squared	0.66	0.00
adj.r.squared	0.65	0.00
sigma	58,196.79	98,553.81
statistic	141.73	—
p.value	0.00	—
df	23.00	—
logLik	−21,473.08	−22,398.09
AIC	42,996.17	44,800.18
BIC	43,132.62	44,811.10
deviance	5,791,540,820,291.45	16,832,374,912,494.90
df.residual	1,710.00	1,733.00
nobs	1,734.00	1,734.00

Forward and Backward Model

```

M0 = lm(Price ~ 1, data = houses) # Null model
Mf = lm(Price ~ . - Test, data = houses) # Full model

```

```

step_back = step(Mf, direction = "backward", trace = TRUE)

```

```

## Start: AIC=38074.34
## Price ~ (Lot.Size + Waterfront + Age + Land.Value + New.Construct +
## Central.Air + Fuel.Type + Heat.Type + Sewer.Type + Living.Area +
## Pct.College + Bedrooms + Fireplaces + Bathrooms + Rooms +
## Test) - Test
##
##           Df Sum of Sq      RSS   AIC
## - Fuel.Type    5 6.3388e+09 5.8081e+12 38066
## - Sewer.Type    2 7.0528e+08 5.8024e+12 38071
## - Fireplaces    1 4.1561e+08 5.8021e+12 38072

```

```

## - Pct.College      1 1.8250e+09 5.8036e+12 38073
## <none>              5.8017e+12 38074
## - Heat.Type        2 2.0965e+10 5.8227e+12 38077
## - Age              1 1.7918e+10 5.8197e+12 38078
## - Central.Air      1 2.7659e+10 5.8294e+12 38081
## - Bedrooms         1 3.1674e+10 5.8334e+12 38082
## - Rooms            1 3.3534e+10 5.8353e+12 38082
## - Lot.Size         1 3.8682e+10 5.8404e+12 38084
## - New.Construct    1 1.3172e+11 5.9335e+12 38111
## - Bathrooms        1 1.5860e+11 5.9603e+12 38119
## - Waterfront       1 2.0282e+11 6.0046e+12 38132
## - Living.Area      1 7.8136e+11 6.5831e+12 38291
## - Land.Value       1 1.2789e+12 7.0806e+12 38418
##
## Step: AIC=38066.23
## Price ~ Lot.Size + Waterfront + Age + Land.Value + New.Construct +
##         Central.Air + Heat.Type + Sewer.Type + Living.Area + Pct.College +
##         Bedrooms + Fireplaces + Bathrooms + Rooms
##
##           Df Sum of Sq      RSS   AIC
## - Sewer.Type      2 2.6524e+08 5.8083e+12 38062
## - Fireplaces      1 4.2121e+08 5.8085e+12 38064
## - Pct.College     1 1.4135e+09 5.8095e+12 38065
## <none>              5.8081e+12 38066
## - Age            1 2.0801e+10 5.8289e+12 38070
## - Central.Air     1 2.7552e+10 5.8356e+12 38072
## - Heat.Type       3 4.2601e+10 5.8507e+12 38073
## - Bedrooms        1 3.0634e+10 5.8387e+12 38073
## - Rooms           1 3.4125e+10 5.8422e+12 38074
## - Lot.Size        1 3.5700e+10 5.8438e+12 38075
## - New.Construct   1 1.3030e+11 5.9384e+12 38103
## - Bathrooms       1 1.5936e+11 5.9674e+12 38111
## - Waterfront      1 2.0083e+11 6.0089e+12 38123
## - Living.Area     1 7.8183e+11 6.5899e+12 38283
## - Land.Value      1 1.2964e+12 7.1044e+12 38414
##
## Step: AIC=38062.31
## Price ~ Lot.Size + Waterfront + Age + Land.Value + New.Construct +
##         Central.Air + Heat.Type + Living.Area + Pct.College + Bedrooms +
##         Fireplaces + Bathrooms + Rooms
##
##           Df Sum of Sq      RSS   AIC
## - Fireplaces      1 4.3961e+08 5.8088e+12 38060
## - Pct.College     1 1.5823e+09 5.8099e+12 38061
## <none>              5.8083e+12 38062
## - Age            1 2.0586e+10 5.8289e+12 38066
## - Central.Air     1 2.7582e+10 5.8359e+12 38069
## - Heat.Type       3 4.2734e+10 5.8511e+12 38069
## - Bedrooms        1 3.0660e+10 5.8390e+12 38069
## - Rooms           1 3.4212e+10 5.8425e+12 38070
## - Lot.Size        1 4.2504e+10 5.8508e+12 38073
## - New.Construct   1 1.3060e+11 5.9389e+12 38099
## - Bathrooms       1 1.6015e+11 5.9685e+12 38107
## - Waterfront      1 2.0168e+11 6.0100e+12 38119

```

```

## - Living.Area      1 7.8292e+11 6.5913e+12 38280
## - Land.Value       1 1.3153e+12 7.1236e+12 38414
##
## Step: AIC=38060.44
## Price ~ Lot.Size + Waterfront + Age + Land.Value + New.Construct +
##       Central.Air + Heat.Type + Living.Area + Pct.College + Bedrooms +
##       Bathrooms + Rooms
##
##              Df  Sum of Sq      RSS    AIC
## - Pct.College   1 1.3964e+09 5.8102e+12 38059
## <none>              5.8088e+12 38060
## - Age           1 2.1035e+10 5.8298e+12 38065
## - Central.Air   1 2.8661e+10 5.8374e+12 38067
## - Heat.Type     3 4.2740e+10 5.8515e+12 38067
## - Bedrooms      1 3.0816e+10 5.8396e+12 38068
## - Rooms         1 3.4012e+10 5.8428e+12 38069
## - Lot.Size      1 4.2744e+10 5.8515e+12 38071
## - New.Construct 1 1.3235e+11 5.9411e+12 38098
## - Bathrooms     1 1.6361e+11 5.9724e+12 38107
## - Waterfront    1 2.0174e+11 6.0105e+12 38118
## - Living.Area   1 8.2677e+11 6.6355e+12 38289
## - Land.Value    1 1.3149e+12 7.1236e+12 38412
##
## Step: AIC=38058.86
## Price ~ Lot.Size + Waterfront + Age + Land.Value + New.Construct +
##       Central.Air + Heat.Type + Living.Area + Bedrooms + Bathrooms +
##       Rooms
##
##              Df  Sum of Sq      RSS    AIC
## <none>              5.8102e+12 38059
## - Age           1 2.0889e+10 5.8311e+12 38063
## - Central.Air   1 2.7338e+10 5.8375e+12 38065
## - Heat.Type     3 4.3183e+10 5.8534e+12 38066
## - Bedrooms      1 3.1701e+10 5.8419e+12 38066
## - Rooms         1 3.4130e+10 5.8443e+12 38067
## - Lot.Size      1 4.4126e+10 5.8543e+12 38070
## - New.Construct 1 1.3183e+11 5.9420e+12 38096
## - Bathrooms     1 1.6327e+11 5.9734e+12 38105
## - Waterfront    1 2.0869e+11 6.0189e+12 38118
## - Living.Area   1 8.2538e+11 6.6356e+12 38287
## - Land.Value    1 1.3439e+12 7.1541e+12 38418

```

```
summary(step_back)
```

```

##
## Call:
## lm(formula = Price ~ Lot.Size + Waterfront + Age + Land.Value +
##     New.Construct + Central.Air + Heat.Type + Living.Area + Bedrooms +
##     Bathrooms + Rooms, data = houses)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -228668  -35010   -4386    27523   456905
##

```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.741e+03  6.546e+03   1.182  0.23717
## Lot.Size      7.372e+03  2.040e+03   3.614  0.00031 ***
## Waterfront1   1.203e+05  1.531e+04   7.860 6.72e-15 ***
## Age          -1.408e+02  5.662e+01  -2.487  0.01299 *
## Land.Value     9.195e-01  4.610e-02  19.946 < 2e-16 ***
## New.Construct1 -4.454e+04  7.131e+03  -6.247 5.26e-10 ***
## Central.Air1   9.639e+03  3.388e+03   2.845  0.00450 **
## Heat.TypeHot Air  9.999e+03  4.032e+03   2.480  0.01324 *
## Heat.TypeHot Water -5.115e+02  5.192e+03  -0.099  0.92154
## Heat.TypeNone  -3.295e+04  2.470e+04  -1.334  0.18240
## Living.Area    7.017e+01  4.489e+00  15.631 < 2e-16 ***
## Bedrooms      -7.798e+03  2.545e+03  -3.063  0.00222 **
## Bathrooms      2.305e+04  3.315e+03   6.952 5.09e-12 ***
## Rooms          3.046e+03  9.583e+02   3.179  0.00151 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58120 on 1720 degrees of freedom
## Multiple R-squared:  0.6548, Adjusted R-squared:  0.6522
## F-statistic: 251 on 13 and 1720 DF, p-value: < 2.2e-16
```

```
step_forward = step(M0, scope = list(lower = M0, upper = Mf), direction = "forward", trace = TRUE)
```

```
## Start: AIC=39877.31
## Price ~ 1
##
##              Df Sum of Sq      RSS   AIC
## + Living.Area  1 8.5687e+12 8.2637e+12 38646
## + Bathrooms    1 6.0140e+12 1.0818e+13 39113
## + Land.Value   1 5.6756e+12 1.1157e+13 39166
## + Rooms        1 4.7877e+12 1.2045e+13 39299
## + Bedrooms     1 2.7271e+12 1.4105e+13 39573
## + Fireplaces   1 2.4250e+12 1.4407e+13 39610
## + Central.Air  1 1.8836e+12 1.4949e+13 39674
## + Fuel.Type    6 1.2364e+12 1.5596e+13 39757
## + Heat.Type    3 1.0882e+12 1.5744e+13 39767
## + Pct.College  1 6.7204e+11 1.6160e+13 39809
## + Age          1 6.6428e+11 1.6168e+13 39809
## + New.Construct 1 4.2546e+11 1.6407e+13 39835
## + Lot.Size     1 4.1693e+11 1.6415e+13 39836
## + Waterfront   1 3.9929e+11 1.6433e+13 39838
## + Sewer.Type   2 1.2253e+11 1.6710e+13 39869
## <none>                1.6832e+13 39877
##
## Step: AIC=38645.67
## Price ~ Living.Area
##
##              Df Sum of Sq      RSS   AIC
## + Land.Value  1 1.5952e+12 6.6684e+12 38276
## + Waterfront  1 4.0797e+11 7.8557e+12 38560
## + Central.Air 1 3.0157e+11 7.9621e+12 38583
## + Bathrooms   1 2.5093e+11 8.0127e+12 38594
```



```

## + Bedrooms      1 1.3104e+11 8.1326e+12 38620
## + Heat.Type     3 1.4920e+11 8.1145e+12 38620
## + Age           1 8.2864e+10 8.1808e+12 38630
## + Pct.College   1 4.4416e+10 8.2192e+12 38638
## + Fuel.Type     6 9.0704e+10 8.1730e+12 38639
## + Fireplaces    1 3.4877e+10 8.2288e+12 38640
## + Lot.Size      1 2.8762e+10 8.2349e+12 38642
## + Sewer.Type    2 3.2585e+10 8.2311e+12 38643
## + New.Construct 1 1.7079e+10 8.2466e+12 38644
## <none>          8.2637e+12 38646
## + Rooms         1 3.0049e+09 8.2607e+12 38647
##
## Step: AIC=38275.76
## Price ~ Living.Area + Land.Value
##
##           Df Sum of Sq      RSS   AIC
## + Bathrooms      1 2.6449e+11 6.4039e+12 38208
## + Waterfront     1 2.1987e+11 6.4485e+12 38220
## + Central.Air    1 1.4573e+11 6.5227e+12 38239
## + Age            1 1.3134e+11 6.5371e+12 38243
## + New.Construct  1 1.0238e+11 6.5660e+12 38251
## + Heat.Type      3 9.4210e+10 6.5742e+12 38257
## + Bedrooms       1 4.9922e+10 6.6185e+12 38265
## + Lot.Size       1 3.4093e+10 6.6343e+12 38269
## + Fireplaces     1 2.8890e+10 6.6395e+12 38270
## <none>           6.6684e+12 38276
## + Rooms          1 6.3530e+09 6.6621e+12 38276
## + Pct.College    1 1.2470e+08 6.6683e+12 38278
## + Sewer.Type     2 7.1046e+08 6.6677e+12 38280
## + Fuel.Type      6 3.0111e+10 6.6383e+12 38280
##
## Step: AIC=38207.58
## Price ~ Living.Area + Land.Value + Bathrooms
##
##           Df Sum of Sq      RSS   AIC
## + Waterfront     1 2.2116e+11 6.1828e+12 38149
## + New.Construct  1 1.2679e+11 6.2771e+12 38175
## + Central.Air    1 7.7307e+10 6.3266e+12 38189
## + Heat.Type      3 7.7237e+10 6.3267e+12 38193
## + Bedrooms       1 4.4322e+10 6.3596e+12 38198
## + Lot.Size       1 4.3886e+10 6.3600e+12 38198
## + Age            1 4.0150e+10 6.3638e+12 38199
## + Fireplaces     1 8.2548e+09 6.3957e+12 38207
## + Rooms          1 8.1022e+09 6.3958e+12 38207
## <none>           6.4039e+12 38208
## + Sewer.Type     2 8.4635e+09 6.3955e+12 38209
## + Pct.College    1 1.4474e+08 6.4038e+12 38210
## + Fuel.Type      6 2.5488e+10 6.3784e+12 38213
##
## Step: AIC=38148.64
## Price ~ Living.Area + Land.Value + Bathrooms + Waterfront
##
##           Df Sum of Sq      RSS   AIC
## + New.Construct  1 1.1335e+11 6.0694e+12 38119

```

```

## + Central.Air      1 8.8033e+10 6.0947e+12 38126
## + Heat.Type        3 7.6625e+10 6.1061e+12 38133
## + Lot.Size         1 4.4526e+10 6.1382e+12 38138
## + Age              1 4.2856e+10 6.1399e+12 38139
## + Bedrooms         1 3.0451e+10 6.1523e+12 38142
## + Rooms            1 1.2782e+10 6.1700e+12 38147
## + Fireplaces       1 9.4172e+09 6.1734e+12 38148
## <none>              6.1828e+12 38149
## + Pct.College      1 1.9722e+09 6.1808e+12 38150
## + Sewer.Type       2 4.4226e+09 6.1784e+12 38151
## + Fuel.Type        6 2.6213e+10 6.1566e+12 38153
##
## Step:  AIC=38118.55
## Price ~ Living.Area + Land.Value + Bathrooms + Waterfront + New.Construct
##
##           Df Sum of Sq      RSS   AIC
## + Heat.Type    3 9.8237e+10 5.9712e+12 38096
## + Central.Air  1 8.0529e+10 5.9889e+12 38097
## + Age          1 7.0091e+10 5.9993e+12 38100
## + Lot.Size     1 3.7391e+10 6.0320e+12 38110
## + Bedrooms     1 3.4099e+10 6.0353e+12 38111
## + Rooms        1 1.3200e+10 6.0562e+12 38117
## <none>          6.0694e+12 38119
## + Fireplaces   1 4.1118e+09 6.0653e+12 38119
## + Pct.College  1 4.0281e+08 6.0690e+12 38120
## + Sewer.Type   2 3.2366e+09 6.0662e+12 38122
## + Fuel.Type    6 2.8410e+10 6.0410e+12 38122
##
## Step:  AIC=38096.26
## Price ~ Living.Area + Land.Value + Bathrooms + Waterfront + New.Construct +
##           Heat.Type
##
##           Df Sum of Sq      RSS   AIC
## + Lot.Size    1 4.5991e+10 5.9252e+12 38085
## + Age         1 3.8214e+10 5.9330e+12 38087
## + Central.Air  1 3.7829e+10 5.9334e+12 38087
## + Bedrooms     1 2.6214e+10 5.9450e+12 38091
## + Rooms        1 1.3479e+10 5.9577e+12 38094
## <none>          5.9712e+12 38096
## + Fireplaces   1 3.0289e+09 5.9682e+12 38097
## + Pct.College  1 7.3319e+08 5.9705e+12 38098
## + Sewer.Type   2 4.0832e+09 5.9671e+12 38099
## + Fuel.Type    5 2.2456e+09 5.9689e+12 38106
##
## Step:  AIC=38084.85
## Price ~ Living.Area + Land.Value + Bathrooms + Waterfront + New.Construct +
##           Heat.Type + Lot.Size
##
##           Df Sum of Sq      RSS   AIC
## + Central.Air  1 4.0496e+10 5.8847e+12 38075
## + Age         1 3.6048e+10 5.8891e+12 38076
## + Bedrooms     1 2.5927e+10 5.8993e+12 38079
## + Rooms        1 1.2355e+10 5.9128e+12 38083
## <none>          5.9252e+12 38085

```

```

## + Fireplaces      1 2.7584e+09 5.9224e+12 38086
## + Pct.College     1 1.0835e+08 5.9251e+12 38087
## + Sewer.Type      2 4.5854e+08 5.9247e+12 38089
## + Fuel.Type       5 9.1477e+09 5.9160e+12 38092
##
## Step: AIC=38074.96
## Price ~ Living.Area + Land.Value + Bathrooms + Waterfront + New.Construct +
##      Heat.Type + Lot.Size + Central.Air
##
##           Df Sum of Sq      RSS   AIC
## + Age      1 2.6085e+10 5.8586e+12 38069
## + Bedrooms 1 1.9374e+10 5.8653e+12 38071
## + Rooms    1 1.4932e+10 5.8698e+12 38073
## <none>                5.8847e+12 38075
## + Pct.College 1 2.0234e+09 5.8827e+12 38076
## + Fireplaces 1 5.6811e+08 5.8841e+12 38077
## + Sewer.Type 2 1.0754e+08 5.8846e+12 38079
## + Fuel.Type  5 7.3644e+09 5.8773e+12 38083
##
## Step: AIC=38069.26
## Price ~ Living.Area + Land.Value + Bathrooms + Waterfront + New.Construct +
##      Heat.Type + Lot.Size + Central.Air + Age
##
##           Df Sum of Sq      RSS   AIC
## + Rooms    1 1.6739e+10 5.8419e+12 38066
## + Bedrooms 1 1.4310e+10 5.8443e+12 38067
## <none>                5.8586e+12 38069
## + Pct.College 1 2.1147e+09 5.8565e+12 38071
## + Fireplaces 1 1.7431e+08 5.8584e+12 38071
## + Sewer.Type 2 4.5136e+08 5.8582e+12 38073
## + Fuel.Type  5 5.1338e+09 5.8535e+12 38078
##
## Step: AIC=38066.3
## Price ~ Living.Area + Land.Value + Bathrooms + Waterfront + New.Construct +
##      Heat.Type + Lot.Size + Central.Air + Age + Rooms
##
##           Df Sum of Sq      RSS   AIC
## + Bedrooms 1 3.1701e+10 5.8102e+12 38059
## <none>                5.8419e+12 38066
## + Pct.College 1 2.2816e+09 5.8396e+12 38068
## + Fireplaces 1 3.2463e+08 5.8415e+12 38068
## + Sewer.Type 2 3.7202e+08 5.8415e+12 38070
## + Fuel.Type  5 4.5496e+09 5.8373e+12 38075
##
## Step: AIC=38058.86
## Price ~ Living.Area + Land.Value + Bathrooms + Waterfront + New.Construct +
##      Heat.Type + Lot.Size + Central.Air + Age + Rooms + Bedrooms
##
##           Df Sum of Sq      RSS   AIC
## <none>                5.8102e+12 38059
## + Pct.College 1 1396417766 5.8088e+12 38060
## + Fireplaces  1 253748311 5.8099e+12 38061
## + Sewer.Type  2 436841631 5.8097e+12 38063
## + Fuel.Type   5 5319314117 5.8049e+12 38067

```

```
summary(step_forward)
```

```
##
## Call:
## lm(formula = Price ~ Living.Area + Land.Value + Bathrooms + Waterfront +
##     New.Construct + Heat.Type + Lot.Size + Central.Air + Age +
##     Rooms + Bedrooms, data = houses)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -228668  -35010   -4386    27523   456905
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.741e+03  6.546e+03   1.182  0.23717
## Living.Area     7.017e+01  4.489e+00  15.631 < 2e-16 ***
## Land.Value      9.195e-01  4.610e-02  19.946 < 2e-16 ***
## Bathrooms      2.305e+04  3.315e+03   6.952 5.09e-12 ***
## Waterfront1     1.203e+05  1.531e+04   7.860 6.72e-15 ***
## New.Construct1  -4.454e+04  7.131e+03  -6.247 5.26e-10 ***
## Heat.TypeHot Air  9.999e+03  4.032e+03   2.480  0.01324 *
## Heat.TypeHot Water -5.115e+02  5.192e+03  -0.099  0.92154
## Heat.TypeNone    -3.295e+04  2.470e+04  -1.334  0.18240
## Lot.Size        7.372e+03  2.040e+03   3.614  0.00031 ***
## Central.Air1     9.639e+03  3.388e+03   2.845  0.00450 **
## Age             -1.408e+02  5.662e+01  -2.487  0.01299 *
## Rooms           3.046e+03  9.583e+02   3.179  0.00151 **
## Bedrooms        -7.798e+03  2.545e+03  -3.063  0.00222 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58120 on 1720 degrees of freedom
## Multiple R-squared:  0.6548, Adjusted R-squared:  0.6522
## F-statistic: 251 on 13 and 1720 DF, p-value: < 2.2e-16
```

Comparison of backward and forward model

```
sjPlot::tab_model(step_forward, step_back, show.ci = FALSE, show.aic = TRUE,
  dv.labels = c("Forward model", "Backward model"))
```

Forward model

Backward model

Predictors

Estimates

p

Estimates

p

(Intercept)
7740.54
0.237
7740.54
0.237
Living Area
70.17
<0.001
70.17
<0.001
Land Value
0.92
<0.001
0.92
<0.001
Bathrooms
23048.18
<0.001
23048.18
<0.001
Waterfront [1]
120327.80
<0.001
120327.80
<0.001
New Construct [1]
-44544.52
<0.001
-44544.52
<0.001
Heat Type [Hot Air]
9998.56
0.013
9998.56
0.013
Heat Type [Hot Water]

-511.48
 0.922
 -511.48
 0.922
 Heat Type [None]
 -32952.33
 0.182
 -32952.33
 0.182
 Lot Size
 7372.04
 <0.001
 7372.04
 <0.001
 Central Air [1]
 9639.20
 0.004
 9639.20
 0.004
 Age
 -140.80
 0.013
 -140.80
 0.013
 Rooms
 3045.91
 0.002
 3045.91
 0.002
 Bedrooms
 -7797.56
 0.002
 -7797.56
 0.002
 Observations
 1734

1734
R2 / R2 adjusted
0.655 / 0.652
0.655 / 0.652
AIC
42981.739
42981.739