Realestate_Yongsan

2023-06-11

데이터 로드 및 변환

```
library(dplyr)
```

```
##
## 다음의 패키지를 부착합니다: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# finalretaildata 불러오기

#setwd("C:₩₩Rtest₩₩realestate")
data_whole <- read.csv("RealEstateData/FinalRetailData_1차 수정.csv", h = T, fileEncoding = "cp
949")
data_whole <- subset(data_whole, Rejion == "용산구")  # 다른 구 분석하려면 이 부분 변경경
summary(data_whole)
```

```
##      index      transaction_id     apartment_id       city
##  Min.   :1706   Min.   :1094671   Min.   :    9   Length:2079
##  1st Qu.:2226   1st Qu.:1095208   1st Qu.: 4953   Class :character
##  Median :2745   Median :1095734   Median : 8541   Mode  :character
##  Mean   :2745   Mean   :1095745   Mean   : 7623
##  3rd Qu.:3264   3rd Qu.:1096273   3rd Qu.:11263
##  Max.   :3784   Max.   :1096876   Max.   :12654
##      dong              jibun              apt              addr_kr
##  Length:2079        Length:2079        Length:2079        Length:2079
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  exclusive_use_area year_of_completion transaction_year_month
##  Min.   : 13.01     Min.   :1969       Length:2079
##  1st Qu.: 65.62     1st Qu.:1997       Class :character
##  Median : 84.98     Median :2001       Mode  :character
##  Mean   :101.61     Mean   :1999
##  3rd Qu.:123.42     3rd Qu.:2007
##  Max.   :244.78     Max.   :2017
##  transaction_date       floor      transaction_real_price      year
##  Length:2079        Min.   :-1.0   Min.   : 13000         Min.   :2017
##  Class :character   1st Qu.: 4.0   1st Qu.: 64000         1st Qu.:2017
##  Mode  :character   Median : 8.0   Median : 82000         Median :2017
##                     Mean   :10.2   Mean   :108215         Mean   :2017
##                     3rd Qu.:15.0   3rd Qu.:113250         3rd Qu.:2017
##                     Max.   :54.0   Max.   :780000         Max.   :2017
##     Latitude        Hardness        Rejion           bigMarket05
##  Min.   :126.9   Min.   :37.52   Length:2079        Min.   :0.0000
##  1st Qu.:127.0   1st Qu.:37.52   Class :character   1st Qu.:0.0000
##  Median :127.0   Median :37.53   Mode  :character   Median :1.0000
##  Mean   :127.0   Mean   :37.53                      Mean   :0.8288
##  3rd Qu.:127.0   3rd Qu.:37.54                      3rd Qu.:2.0000
##  Max.   :127.0   Max.   :37.55                      Max.   :3.0000
##   bigMarket10     bigMarket15       school05        school10
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   : 0.000
##  1st Qu.:1.000   1st Qu.:3.000   1st Qu.:1.000   1st Qu.: 3.000
##  Median :2.000   Median :4.000   Median :2.000   Median : 5.000
##  Mean   :2.082   Mean   :4.165   Mean   :2.019   Mean   : 5.598
##  3rd Qu.:3.000   3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.: 8.000
##  Max.   :5.000   Max.   :9.000   Max.   :6.000   Max.   :20.000
##     school15        subway05         subway10        subway15
##  Min.   : 5.00   Min.   :0.0000   Min.   :0.000   Min.   : 1.000
##  1st Qu.: 6.00   1st Qu.:0.0000   1st Qu.:2.000   1st Qu.: 5.000
##  Median : 8.00   Median :1.0000   Median :3.000   Median : 7.000
##  Mean   :11.95   Mean   :0.9865   Mean   :3.881   Mean   : 7.268
##  3rd Qu.:19.50   3rd Qu.:2.0000   3rd Qu.:6.000   3rd Qu.:10.000
##  Max.   :32.00   Max.   :3.0000   Max.   :9.000   Max.   :16.000
##    hospital05      hospital10       hospital15       movie05
##  Min.   : 0.00   Min.   :  2.00   Min.   : 33.0   Min.   : 0.000
##  1st Qu.:12.00   1st Qu.: 45.00   1st Qu.: 97.0   1st Qu.: 1.000
##  Median :16.00   Median : 80.00   Median :106.0   Median : 3.000
##  Mean   :20.55   Mean   : 74.45   Mean   :133.1   Mean   : 4.721
##  3rd Qu.:32.00   3rd Qu.: 94.00   3rd Qu.:168.0   3rd Qu.: 5.000
```

```
##  Max.   :48.00   Max.   :189.00   Max.   :341.0   Max.   :34.000
##     movie10         movie15          kid05           kid10
##  Min.   : 4.00   Min.   : 14.00   Min.   : 2.00   Min.   : 7.00
##  1st Qu.:12.00   1st Qu.: 21.00   1st Qu.: 6.00   1st Qu.:18.00
##  Median :15.00   Median : 29.00   Median :10.00   Median :25.00
##  Mean   :21.97   Mean   : 38.98   Mean   :10.59   Mean   :26.74
##  3rd Qu.:31.00   3rd Qu.: 49.00   3rd Qu.:13.00   3rd Qu.:28.00
##  Max.   :71.00   Max.   :101.00   Max.   :29.00   Max.   :68.00
##      kid15          office05        office10        office15
##  Min.   : 25.00   Min.   : 0.000   Min.   : 2.00   Min.   : 5.00
##  1st Qu.: 35.00   1st Qu.: 1.000   1st Qu.: 5.00   1st Qu.: 9.00
##  Median : 43.00   Median : 2.000   Median : 9.00   Median :18.00
##  Mean   : 52.42   Mean   : 4.046   Mean   :13.37   Mean   :25.82
##  3rd Qu.: 67.00   3rd Qu.: 4.000   3rd Qu.:14.00   3rd Qu.:28.00
##  Max.   :124.00   Max.   :22.000   Max.   :70.00   Max.   :79.00
```

```
str(data_whole)
```

```
## 'data.frame':    2079 obs. of  39 variables:
##  $ index                : int  1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 ...
##  $ transaction_id       : int  1094671 1094672 1094674 1094675 1094676 1094677 1094678 1094
679 1094680 1094681 ...
##  $ apartment_id         : int  12561 5299 8539 6714 4070 4070 393 5425 12556 12556 ...
##  $ city                 : chr  "서울특별시" "서울특별시" "서울특별시" "서울특별시" ...
##  $ dong                 : chr  "후암동" "후암동" "원효로1가" "신창동" ...
##  $ jibun                : chr  "423-1" "458" "41" "102" ...
##  $ apt                  : chr  "후암미주" "브라운스톤남산" "용산 더프라임" "세방리버하이빌"
...
##  $ addr_kr              : chr  "후암동 423-1 후암미주" "후암동 458 브라운스톤남산" "원효로1
가 41 용산 더프라임" "신창동 102 세방리버하이빌" ...
##  $ exclusive_use_area   : num  62.3 166.6 46 84.5 85 ...
##  $ year_of_completion   : int  1980 2004 2014 2005 2001 2001 2005 1977 2010 2010 ...
##  $ transaction_year_month: chr  "2017-01-01" "2017-01-01" "2017-01-01" "2017-01-01" ...
##  $ transaction_date     : chr  "1~10" "11~20" "21~31" "11~20" ...
##  $ floor                : int  6 4 3 8 5 16 9 11 17 9 ...
##  $ transaction_real_price: int  49000 112000 53000 52500 59000 57000 173000 73000 62500 6350
0 ...
##  $ year                 : int  2017 2017 2017 2017 2017 2017 2017 2017 2017 2017 ...
##  $ Latitude             : num  127 127 127 127 127 ...
##  $ Hardness             : num  37.6 37.6 37.5 37.5 37.5 ...
##  $ Rejion               : chr  "용산구" "용산구" "용산구" "용산구" ...
##  $ bigMarket05          : int  0 0 1 1 0 0 0 0 1 1 ...
##  $ bigMarket10          : int  2 2 2 3 2 2 3 1 4 4 ...
##  $ bigMarket15          : int  2 2 4 7 6 6 5 6 8 8 ...
##  $ school05             : int  0 1 5 4 4 4 2 3 3 3 ...
##  $ school10             : int  7 6 14 8 5 5 5 4 12 12 ...
##  $ school15             : int  28 28 21 17 12 12 13 10 26 26 ...
##  $ subway05             : int  1 1 1 0 0 0 0 0 3 3 ...
##  $ subway10             : int  6 6 6 7 3 3 3 1 6 6 ...
##  $ subway15             : int  12 10 11 10 9 9 8 9 13 13 ...
##  $ hospital05           : int  15 14 39 11 13 13 14 6 20 20 ...
##  $ hospital10           : int  89 86 81 160 117 117 107 85 182 182 ...
##  $ hospital15           : int  266 256 158 245 203 203 202 209 304 304 ...
##  $ movie05              : int  8 8 4 3 1 1 2 2 6 6 ...
##  $ movie10              : int  27 23 27 12 9 9 6 7 18 18 ...
##  $ movie15              : int  89 87 49 25 21 21 17 19 32 32 ...
##  $ kid05                : int  11 11 6 24 16 16 13 10 21 21 ...
##  $ kid10                : int  29 28 27 47 41 41 40 36 66 66 ...
##  $ kid15                : int  57 57 70 93 81 81 73 72 113 113 ...
##  $ office05             : int  8 8 5 2 1 1 1 2 3 3 ...
##  $ office10             : int  22 21 13 12 9 9 10 7 15 15 ...
##  $ office15             : int  68 62 21 22 18 18 16 16 32 32 ...
```

```
data_whole %>% colnames()
```

```
##  [1] "index"                "transaction_id"         "apartment_id"
##  [4] "city"                 "dong"                   "jibun"
##  [7] "apt"                  "addr_kr"                "exclusive_use_area"
## [10] "year_of_completion"   "transaction_year_month" "transaction_date"
## [13] "floor"                "transaction_real_price" "year"
## [16] "Latitude"             "Hardness"               "Rejion"
## [19] "bigMarket05"          "bigMarket10"            "bigMarket15"
## [22] "school05"             "school10"               "school15"
## [25] "subway05"             "subway10"               "subway15"
## [28] "hospital05"           "hospital10"             "hospital15"
## [31] "movie05"              "movie10"                "movie15"
## [34] "kid05"                "kid10"                  "kid15"
## [37] "office05"             "office10"               "office15"
```

```
filterCol<-c("index", "transaction_id", "apartment_id", "city", "jibun", "apt", "addr_kr", "Lat
itude", "Hardness", "year", "Rejion")
data_whole<-data_whole %>% select(-all_of(filterCol))
str(data_whole)
```

```
## 'data.frame':    2079 obs. of  28 variables:
##  $ dong                  : chr  "후암동" "후암동" "원효로1가" "신창동" ...
##  $ exclusive_use_area    : num  62.3 166.6 46 84.5 85 ...
##  $ year_of_completion    : int  1980 2004 2014 2005 2001 2001 2005 1977 2010 2010 ...
##  $ transaction_year_month: chr  "2017-01-01" "2017-01-01" "2017-01-01" "2017-01-01" ...
##  $ transaction_date      : chr  "1~10" "11~20" "21~31" "11~20" ...
##  $ floor                 : int  6 4 3 8 5 16 9 11 17 9 ...
##  $ transaction_real_price: int  49000 112000 53000 52500 59000 57000 173000 73000 62500 6350
0 ...
##  $ bigMarket05           : int  0 0 1 1 0 0 0 0 1 1 ...
##  $ bigMarket10           : int  2 2 2 3 2 2 3 1 4 4 ...
##  $ bigMarket15           : int  2 2 4 7 6 6 5 6 8 8 ...
##  $ school05              : int  0 1 5 4 4 4 2 3 3 3 ...
##  $ school10              : int  7 6 14 8 5 5 5 4 12 12 ...
##  $ school15              : int  28 28 21 17 12 12 13 10 26 26 ...
##  $ subway05              : int  1 1 1 0 0 0 0 0 3 3 ...
##  $ subway10              : int  6 6 6 7 3 3 3 1 6 6 ...
##  $ subway15              : int  12 10 11 10 9 9 8 9 13 13 ...
##  $ hospital05            : int  15 14 39 11 13 13 14 6 20 20 ...
##  $ hospital10            : int  89 86 81 160 117 117 107 85 182 182 ...
##  $ hospital15            : int  266 256 158 245 203 203 202 209 304 304 ...
##  $ movie05               : int  8 8 4 3 1 1 2 2 6 6 ...
##  $ movie10               : int  27 23 27 12 9 9 6 7 18 18 ...
##  $ movie15               : int  89 87 49 25 21 21 17 19 32 32 ...
##  $ kid05                 : int  11 11 6 24 16 16 13 10 21 21 ...
##  $ kid10                 : int  29 28 27 47 41 41 40 36 66 66 ...
##  $ kid15                 : int  57 57 70 93 81 81 73 72 113 113 ...
##  $ office05              : int  8 8 5 2 1 1 1 2 3 3 ...
##  $ office10              : int  22 21 13 12 9 9 10 7 15 15 ...
##  $ office15              : int  68 62 21 22 18 18 16 16 32 32 ...
```

```
# 면적당 가격 변수 추가 및 real_price 변수 제거
data_whole$transaction_real_price <- as.numeric(data_whole$transaction_real_price)
data_whole$unit_price <- data_whole$transaction_real_price / data_whole$exclusive_use_area
data_whole$transaction_real_price <- NULL
str(data_whole)
```

```
## 'data.frame':    2079 obs. of  28 variables:
##  $ dong                  : chr  "후암동" "후암동" "원효로1가" "신창동" ...
##  $ exclusive_use_area    : num  62.3 166.6 46 84.5 85 ...
##  $ year_of_completion    : int  1980 2004 2014 2005 2001 2001 2005 1977 2010 2010 ...
##  $ transaction_year_month: chr  "2017-01-01" "2017-01-01" "2017-01-01" "2017-01-01" ...
##  $ transaction_date      : chr  "1~10" "11~20" "21~31" "11~20" ...
##  $ floor                 : int  6 4 3 8 5 16 9 11 17 9 ...
##  $ bigMarket05           : int  0 0 1 1 0 0 0 0 1 1 ...
##  $ bigMarket10           : int  2 2 2 3 2 2 3 1 4 4 ...
##  $ bigMarket15           : int  2 2 4 7 6 6 5 6 8 8 ...
##  $ school05              : int  0 1 5 4 4 4 2 3 3 3 ...
##  $ school10              : int  7 6 14 8 5 5 5 4 12 12 ...
##  $ school15              : int  28 28 21 17 12 12 13 10 26 26 ...
##  $ subway05              : int  1 1 1 0 0 0 0 0 3 3 ...
##  $ subway10              : int  6 6 6 7 3 3 3 1 6 6 ...
##  $ subway15              : int  12 10 11 10 9 9 8 9 13 13 ...
##  $ hospital05            : int  15 14 39 11 13 13 14 6 20 20 ...
##  $ hospital10            : int  89 86 81 160 117 117 107 85 182 182 ...
##  $ hospital15            : int  266 256 158 245 203 203 202 209 304 304 ...
##  $ movie05               : int  8 8 4 3 1 1 2 2 6 6 ...
##  $ movie10               : int  27 23 27 12 9 9 6 7 18 18 ...
##  $ movie15               : int  89 87 49 25 21 21 17 19 32 32 ...
##  $ kid05                 : int  11 11 6 24 16 16 13 10 21 21 ...
##  $ kid10                 : int  29 28 27 47 41 41 40 36 66 66 ...
##  $ kid15                 : int  57 57 70 93 81 81 73 72 113 113 ...
##  $ office05              : int  8 8 5 2 1 1 1 2 3 3 ...
##  $ office10              : int  22 21 13 12 9 9 10 7 15 15 ...
##  $ office15              : int  68 62 21 22 18 18 16 16 32 32 ...
##  $ unit_price            : num  787 672 1152 621 694 ...
```

```
# transaction_month 변수 추가 및 transaction_year_month, transaction_date, apt 변수 제거
data_whole$transaction_month <- substr(data_whole$transaction_year_month, 6, 7)
data_whole$transaction_year_month <- NULL
data_whole$transaction_date <- NULL
data_whole$apt <- NULL
str(data_whole)
```

```
## 'data.frame':    2079 obs. of  27 variables:
##  $ dong              : chr  "후암동" "후암동" "원효로1가" "신창동" ...
##  $ exclusive_use_area: num  62.3 166.6 46 84.5 85 ...
##  $ year_of_completion: int  1980 2004 2014 2005 2001 2001 2005 1977 2010 2010 ...
##  $ floor             : int  6 4 3 8 5 16 9 11 17 9 ...
##  $ bigMarket05       : int  0 0 1 1 0 0 0 0 1 1 ...
##  $ bigMarket10       : int  2 2 2 3 2 2 3 1 4 4 ...
##  $ bigMarket15       : int  2 2 4 7 6 6 5 6 8 8 ...
##  $ school05          : int  0 1 5 4 4 4 2 3 3 3 ...
##  $ school10          : int  7 6 14 8 5 5 5 4 12 12 ...
##  $ school15          : int  28 28 21 17 12 12 13 10 26 26 ...
##  $ subway05          : int  1 1 1 0 0 0 0 0 3 3 ...
##  $ subway10          : int  6 6 6 7 3 3 3 1 6 6 ...
##  $ subway15          : int  12 10 11 10 9 9 8 9 13 13 ...
##  $ hospital05        : int  15 14 39 11 13 13 14 6 20 20 ...
##  $ hospital10        : int  89 86 81 160 117 117 107 85 182 182 ...
##  $ hospital15        : int  266 256 158 245 203 203 202 209 304 304 ...
##  $ movie05           : int  8 8 4 3 1 1 2 2 6 6 ...
##  $ movie10           : int  27 23 27 12 9 9 6 7 18 18 ...
##  $ movie15           : int  89 87 49 25 21 21 17 19 32 32 ...
##  $ kid05             : int  11 11 6 24 16 16 13 10 21 21 ...
##  $ kid10             : int  29 28 27 47 41 41 40 36 66 66 ...
##  $ kid15             : int  57 57 70 93 81 81 73 72 113 113 ...
##  $ office05          : int  8 8 5 2 1 1 1 2 3 3 ...
##  $ office10          : int  22 21 13 12 9 9 10 7 15 15 ...
##  $ office15          : int  68 62 21 22 18 18 16 16 32 32 ...
##  $ unit_price        : num  787 672 1152 621 694 ...
##  $ transaction_month : chr  "01" "01" "01" "01" ...
```

```
#주성동은 예측할 경우 오류가 있어 서빙고동으로 변환
data_whole$dong <- ifelse (data_whole$dong == "주성동" , "서빙고동", data_whole$dong)

# factor 형으로 변환
data_whole$year <- as.factor(data_whole$year)
data_whole$dong <- as.factor(data_whole$dong)
data_whole$transaction_month <- as.factor(data_whole$transaction_month) # 거래월에 따른 가격 변화 확인

# 변환 결과 확인
str(data_whole)
```

```
## 'data.frame':    2079 obs. of  28 variables:
##  $ dong              : Factor w/ 24 levels "도원동","동빙고동",..: 24 24 13 9 6 6 18 15 23 2
3 ...
##  $ exclusive_use_area: num  62.3 166.6 46 84.5 85 ...
##  $ year_of_completion: int  1980 2004 2014 2005 2001 2001 2005 1977 2010 2010 ...
##  $ floor             : int  6 4 3 8 5 16 9 11 17 9 ...
##  $ bigMarket05       : int  0 0 1 1 0 0 0 0 1 1 ...
##  $ bigMarket10       : int  2 2 2 3 2 2 3 1 4 4 ...
##  $ bigMarket15       : int  2 2 4 7 6 6 5 6 8 8 ...
##  $ school05          : int  0 1 5 4 4 4 2 3 3 3 ...
##  $ school10          : int  7 6 14 8 5 5 5 4 12 12 ...
##  $ school15          : int  28 28 21 17 12 12 13 10 26 26 ...
##  $ subway05          : int  1 1 1 0 0 0 0 0 3 3 ...
##  $ subway10          : int  6 6 6 7 3 3 3 1 6 6 ...
##  $ subway15          : int  12 10 11 10 9 9 8 9 13 13 ...
##  $ hospital05        : int  15 14 39 11 13 13 14 6 20 20 ...
##  $ hospital10        : int  89 86 81 160 117 117 107 85 182 182 ...
##  $ hospital15        : int  266 256 158 245 203 203 202 209 304 304 ...
##  $ movie05           : int  8 8 4 3 1 1 2 2 6 6 ...
##  $ movie10           : int  27 23 27 12 9 9 6 7 18 18 ...
##  $ movie15           : int  89 87 49 25 21 21 17 19 32 32 ...
##  $ kid05             : int  11 11 6 24 16 16 13 10 21 21 ...
##  $ kid10             : int  29 28 27 47 41 41 40 36 66 66 ...
##  $ kid15             : int  57 57 70 93 81 81 73 72 113 113 ...
##  $ office05          : int  8 8 5 2 1 1 1 2 3 3 ...
##  $ office10          : int  22 21 13 12 9 9 10 7 15 15 ...
##  $ office15          : int  68 62 21 22 18 18 16 16 32 32 ...
##  $ unit_price        : num  787 672 1152 621 694 ...
##  $ transaction_month : Factor w/ 11 levels "01","02","03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year              : Factor w/ 42 levels "1969","1970",..: 11 29 39 30 26 26 30 9 35 35
...
```

컬럼 값 Exploration 및 데이터 변환

```
library(ggplot2)

# year of completion -- 준공년도
summary(data_whole$year_of_completion)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1969    1997    2001    1999    2007    2017
```

```
data_whole$year_of_completion_f <- cut(data_whole$year_of_completion, breaks = c(0, 1997, 2001,
2007, Inf), labels = c("1st", "2nd", "3rd", "4th"))
data_whole$year_of_completion <- NULL
summary(data_whole$year_of_completion_f)
```

```
## 1st 2nd 3rd 4th
## 536 603 461 479
```

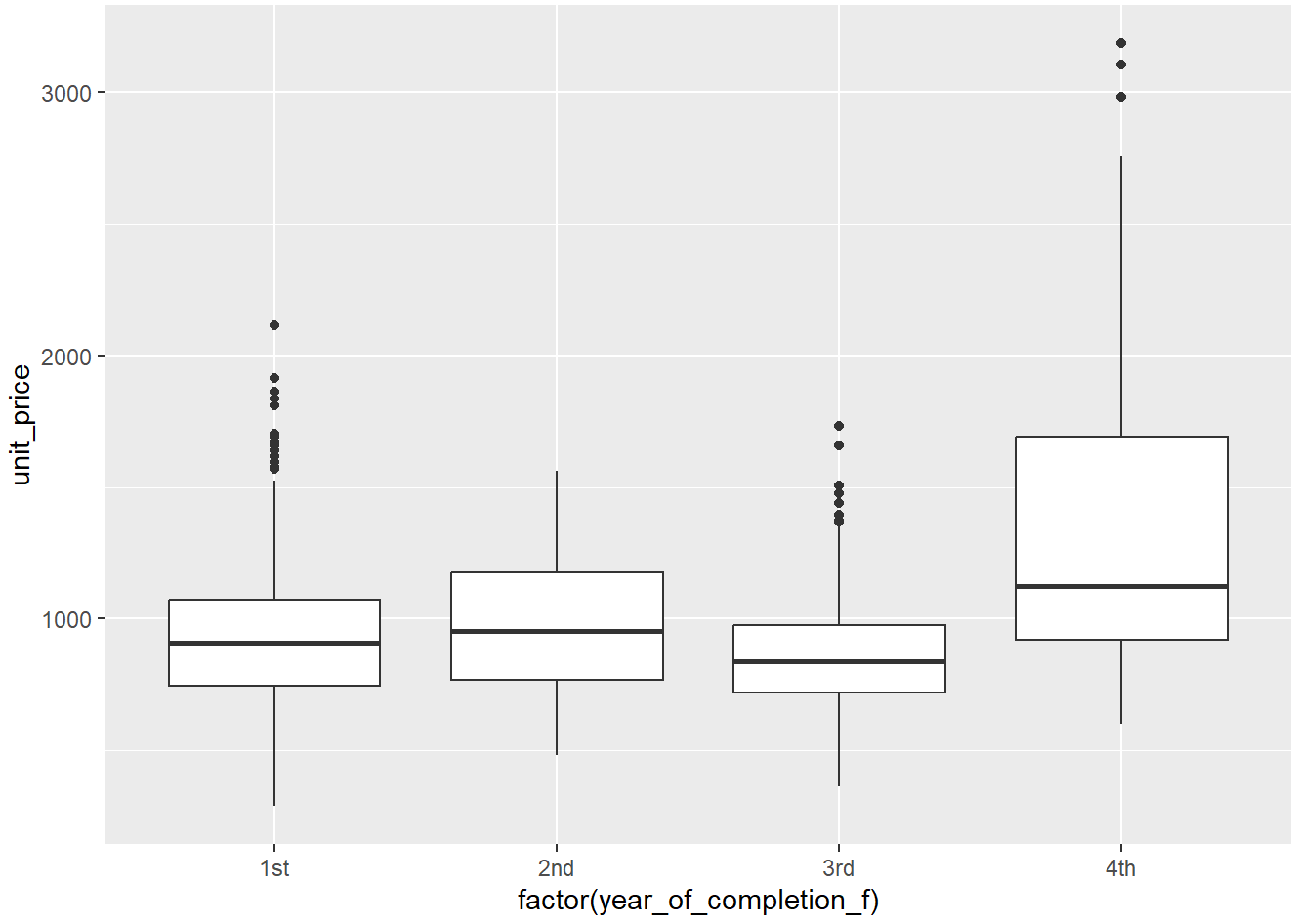
```
# 전체 가격 분포
data_whole %>% summarize(count = n(), avg_price = mean(unit_price), std_price = sd(unit_price))
```

```
##   count avg_price std_price
## 1  2079  1008.265  347.2092
```

```
# 준공년도 factor별 가격 분포
data_whole %>% group_by(year_of_completion_f) %>%
  summarize(count = n(), avg_price = mean(unit_price), std_price = sd(unit_price))
```

```
## # A tibble: 4 × 4
##   year_of_completion_f count avg_price std_price
##   <fct>                <int>     <dbl>     <dbl>
## 1 1st                    536      923.      277.
## 2 2nd                    603      986.      244.
## 3 3rd                    461      863.      196.
## 4 4th                    479     1272.      477.
```

```
ggplot(data = data_whole, aes(x = factor(year_of_completion_f), y = unit_price)) + geom_boxplot
()
```




```
# 동별 가격 분포
summary(data_whole$dong)
```

```
##    도원동  동빙고동    동자동    문배동    보광동    산천동  서빙고동    신계동
##       100         4        18       160        41       103        64        81
##    신창동    용문동 용산동2가 용산동5가 원효로1가 원효로2가 원효로4가    이촌동
##        12        18         8        42        81         1        64       696
##  이태원동    청암동 한강로1가 한강로2가 한강로3가    한남동    효창동    후암동
##        66         8        29        52        77       274        53        27
```

```
data_whole %>% group_by(dong) %>%
  summarize(count = n(), avg_price = mean(unit_price), std_price = sd(unit_price)) # dong별 평
균 및 표준편차
```

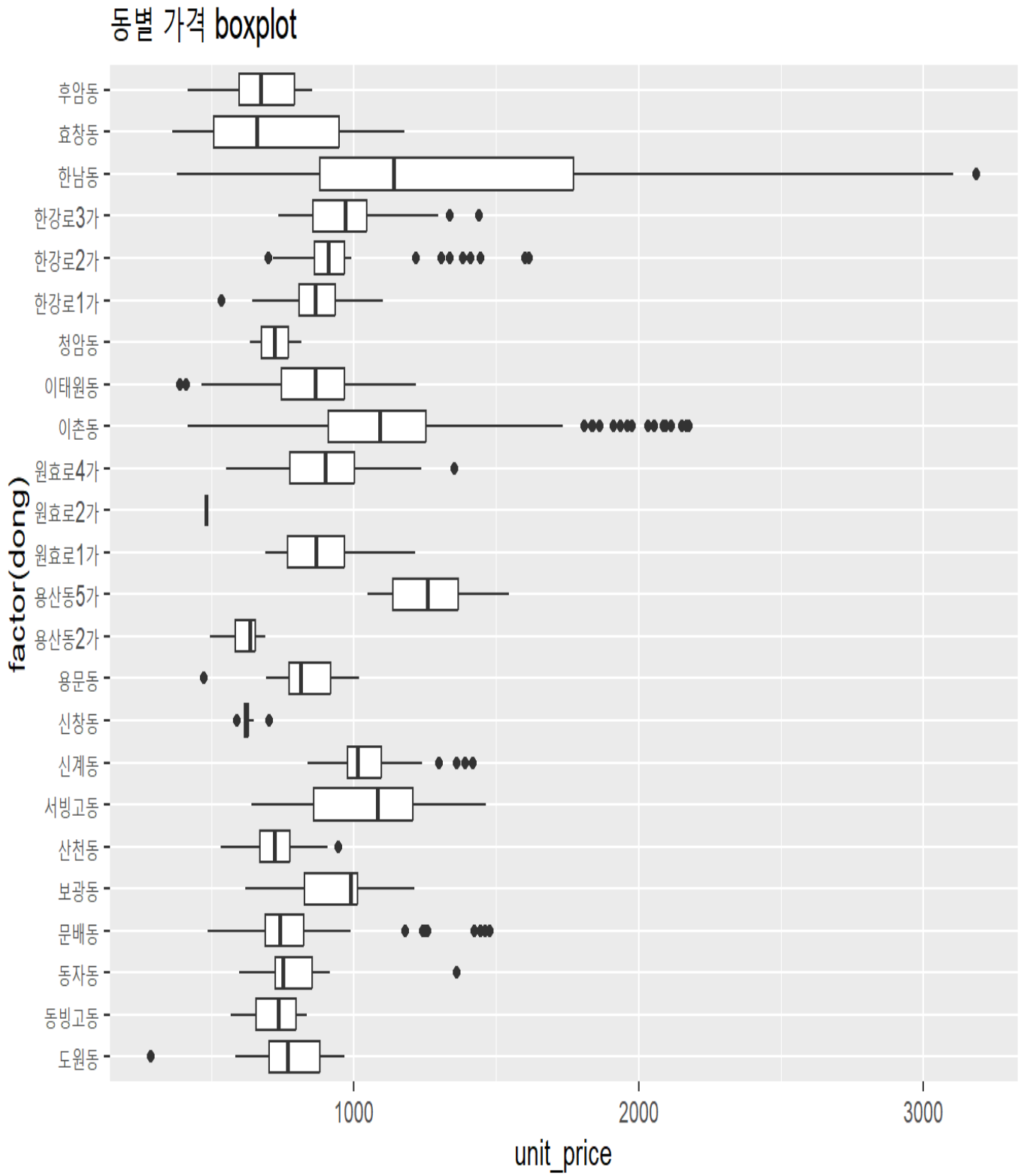
```
## # A tibble: 24 × 4
##    dong     count avg_price std_price
##    <fct>    <int>     <dbl>     <dbl>
##  1 도원동     100      780.     116.
##  2 동빙고동     4      718.     118.
##  3 동자동      18      800.     164.
##  4 문배동     160      804.     192.
##  5 보광동      41      944.     148.
##  6 산천동     103      728.      81.6
##  7 서빙고동    64     1059.     208.
##  8 신계동      81     1038.     117.
##  9 신창동      12      629.      27.5
## 10 용문동      18      829.     131.
## # ℹ 14 more rows
```

```
View(data_whole %>% group_by(dong) %>%
  summarize(count = n(), avg_price = mean(unit_price), std_price = sd(unit_price))) # dong별 평
균 및 표준편차
```

```
# unit price
summary(data_whole$unit_price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   287.2   768.9   942.4  1008.3  1147.5  3186.5
```

```
ggplot(data = data_whole, aes(x = factor(dong), y = unit_price)) + geom_boxplot() + coord_flip
() +ggtitle("동별 가격 boxplot")
```



용산구 단위당 가격 분석

트레이닝 데이터와 테스트 데이터로 split

```
# Data transformation for Tree & Regression Model
data_whole1 <- data_whole

install.packages('caTools', repos ="http://cran.us.r-project.org")
```

```
## 패키지 'caTools'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:₩Users₩LUISW₩AppData₩Local₩Temp₩Rtmpkp5M2w₩downloaded_packages
```

```
library(caTools)

set.seed(123)
sample = sample.split(data_whole1$unit_price, SplitRatio = .7)
data_train1 = subset(data_whole1, sample == TRUE)
data_test1  = subset(data_whole1, sample == FALSE)

str(data_train1)
```

```
## 'data.frame':    1455 obs. of  28 variables:
##  $ dong               : Factor w/ 24 levels "도원동","동빙고동",..: 24 13 6 18 23 23 23 23
1 10 ...
##  $ exclusive_use_area  : num  62.3 46 85 223.8 59.4 ...
##  $ floor               : int  6 3 16 9 17 9 9 6 1 6 ...
##  $ bigMarket05         : int  0 1 0 0 1 1 1 0 0 0 ...
##  $ bigMarket10         : int  2 2 2 3 4 4 4 4 5 5 ...
##  $ bigMarket15         : int  2 4 6 5 8 8 8 8 8 7 ...
##  $ school05            : int  0 5 4 2 3 3 3 5 3 2 ...
##  $ school10            : int  7 14 5 5 12 12 13 14 11 10 ...
##  $ school15            : int  28 21 12 13 26 26 28 29 24 19 ...
##  $ subway05            : int  1 1 0 0 3 3 3 2 2 2 ...
##  $ subway10            : int  6 6 3 3 6 6 6 8 6 7 ...
##  $ subway15            : int  12 11 9 8 13 13 13 15 14 14 ...
##  $ hospital05          : int  15 39 13 14 20 20 35 14 24 31 ...
##  $ hospital10          : int  89 81 117 107 182 182 189 132 169 98 ...
##  $ hospital15          : int  266 158 203 202 304 304 296 299 316 293 ...
##  $ movie05             : int  8 4 1 2 6 6 6 8 5 6 ...
##  $ movie10             : int  27 27 9 6 18 18 18 18 17 17 ...
##  $ movie15             : int  89 49 21 17 32 32 35 42 36 35 ...
##  $ kid05               : int  11 6 16 13 21 21 22 13 18 18 ...
##  $ kid10               : int  29 27 41 40 66 66 66 59 57 53 ...
##  $ kid15               : int  57 70 81 73 113 113 115 102 100 93 ...
##  $ office05            : int  8 5 1 1 3 3 3 3 2 1 ...
##  $ office10            : int  22 13 9 10 15 15 15 13 13 12 ...
##  $ office15            : int  68 21 18 16 32 32 31 33 32 27 ...
##  $ unit_price          : num  787 1152 671 773 1052 ...
##  $ transaction_month   : Factor w/ 11 levels "01","02","03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year                : Factor w/ 42 levels "1969","1970",..: 11 39 26 30 35 35 31 1 3 36
...
##  $ year_of_completion_f: Factor w/ 4 levels "1st","2nd","3rd",..: 1 4 2 3 4 4 3 1 1 4 ...
```

```
str(data_test1)
```

```
## 'data.frame':    624 obs. of  28 variables:
##  $ dong                : Factor w/ 24 levels "도원동","동빙고동",..: 24 9 6 15 23 4 4 4 8 8
...
##  $ exclusive_use_area  : num  166.6 84.5 85 78.1 59.4 ...
##  $ floor               : int  4 8 5 11 3 7 21 16 19 1 ...
##  $ bigMarket05         : int  0 1 0 0 1 1 1 1 2 2 ...
##  $ bigMarket10         : int  2 3 2 1 4 2 3 3 3 3 ...
##  $ bigMarket15         : int  2 7 6 6 8 5 5 4 6 6 ...
##  $ school05            : int  1 4 4 3 3 6 2 2 1 1 ...
##  $ school10            : int  6 8 5 4 12 11 9 8 10 10 ...
##  $ school15            : int  28 17 12 10 26 21 23 22 22 22 ...
##  $ subway05            : int  1 0 0 0 3 1 1 3 1 1 ...
##  $ subway10            : int  6 7 3 1 6 6 9 9 8 8 ...
##  $ subway15            : int  10 10 9 9 13 9 9 9 12 12 ...
##  $ hospital05          : int  14 11 13 6 20 26 15 14 15 15 ...
##  $ hospital10          : int  86 160 117 85 182 91 98 107 107 107 ...
##  $ hospital15          : int  256 245 203 209 304 153 151 142 184 184 ...
##  $ movie05             : int  8 3 1 2 6 3 3 3 1 1 ...
##  $ movie10             : int  23 12 9 7 18 28 35 31 33 33 ...
##  $ movie15             : int  87 25 21 19 32 46 44 44 45 45 ...
##  $ kid05               : int  11 24 16 10 21 10 12 10 11 11 ...
##  $ kid10               : int  28 47 41 36 66 22 28 24 33 33 ...
##  $ kid15               : int  57 93 81 72 113 65 64 59 71 71 ...
##  $ office05            : int  8 2 1 2 3 3 5 5 2 2 ...
##  $ office10            : int  21 12 9 7 15 13 12 12 12 12 ...
##  $ office15            : int  62 22 18 16 32 19 18 16 19 19 ...
##  $ unit_price          : num  672 621 694 935 1052 ...
##  $ transaction_month   : Factor w/ 11 levels "01","02","03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year                : Factor w/ 42 levels "1969","1970",..: 29 30 26 9 35 31 40 37 36 36
...
##  $ year_of_completion_f: Factor w/ 4 levels "1st","2nd","3rd",..: 3 3 2 1 4 3 4 4 4 4 ...
```

```
mean(data_train1$unit_price)
```

```
## [1] 1004.898
```

```
mean(data_test1$unit_price)
```

```
## [1] 1016.117
```

Decision Tree

```
install.packages("rpart", repos ="http://cran.us.r-project.org")
```

```
## 패키지 'rpart'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
```

```
## Warning: 패키지 'rpart'의 이전설치를 삭제할 수 없습니다
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE):
## C:WUsersWLUISWWorkSpaceWRWR-4.3.0WlibraryW00L0CKWrpartWlibsWx64Wrpart.dll를
## C:WUsersWLUISWWorkSpaceWRWR-4.3.0WlibraryWrpartWlibsWx64Wrpart.dll로 복사하는데
## 문제가 발생했습니다: Permission denied
```

```
## Warning: 'rpart'를 복구하였습니다
```

```
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:WUsersWLUISWAppDataWLocalWTempWRtmpkp5M2wWdownloaded_packages
```

```
library(rpart)
tree1 <- rpart(unit_price~.-year,
               data=data_train1,
               method = "anova",
               control = rpart.control(minsplit = 50, maxdepth = 5))

# tree 결과
print(tree1)
```

```
## n= 1455
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 1455 171065500.00 1004.8980
##    2) movie05< 18.5 1358  93510750.00  953.5836
##      4) office15>=9.5 967  37659330.00  868.8085
##        8) dong=도원동,동빙고동,동자동,문배동,보광동,산천동,신창동,용문동,용산동2가,원효로1
가,원효로2가,원효로4가,이촌동,이태원동,청암동,한강로1가,한남동,효창동,후암동 763  23819710.00
823.5971
##         16) exclusive_use_area>=50.20805 694  16313030.00  802.7150
##           32) dong=도원동,동빙고동,동자동,문배동,산천동,신창동,용산동2가,청암동,효창동,후암
동 330   4951899.00  735.8568 *
##           33) dong=보광동,용문동,원효로1가,원효로4가,이촌동,이태원동,한강로1가,한남동 364
8548702.00  863.3283 *
##         17) exclusive_use_area< 50.20805 69   4160236.00 1033.6290 *
##        9) dong=서빙고동,신계동,용산동5가,한강로2가,한강로3가 204   6446682.00 1037.9080
##         18) year_of_completion_f=2nd,3rd 89   1603879.00  924.2603 *
##         19) year_of_completion_f=1st,4th 115   2803673.00 1125.8620 *
##      5) office15< 9.5 391  31714310.00 1163.2450
##       10) kid15>=29.5 353  17464460.00 1107.8880
##         20) year_of_completion_f=1st 93   4851155.00  899.4833
##           40) exclusive_use_area>=146.465 17     44235.33  586.3046 *
##           41) exclusive_use_area< 146.465 76   2766579.00  969.5364 *
##         21) year_of_completion_f=2nd,3rd,4th 260   7129286.00 1182.4330
##           42) exclusive_use_area>=71.805 162   3498333.00 1113.8260 *
##           43) exclusive_use_area< 71.805 98   1607983.00 1295.8430 *
##       11) kid15< 29.5 38   3119511.00 1677.4790 *
##    3) movie05>=18.5 97  23917410.00 1723.2970
##      6) exclusive_use_area< 193.121 21   5601778.00 1146.5890 *
##      7) exclusive_use_area>=193.121 76   9401291.00 1882.6510 *
```

```
summary(tree1)
```

```
## Call:
## rpart(formula = unit_price ~ . - year, data = data_train1, method = "anova",
##     control = rpart.control(minsplit = 50, maxdepth = 5))
##   n= 1455
##
##            CP nsplit rel error    xerror       xstd
## 1  0.31354846      0 1.0000000 1.0024672 0.06632284
## 2  0.14109867      1 0.6864515 0.6897144 0.04020517
## 3  0.06506483      2 0.5453529 0.5620399 0.03676105
## 4  0.05211072      3 0.4802880 0.5332132 0.03582402
## 5  0.04321702      4 0.4281773 0.4871527 0.02929395
## 6  0.03205798      5 0.3849603 0.4114685 0.02558263
## 7  0.01956231      6 0.3529023 0.3706092 0.02429658
## 8  0.01644067      7 0.3333400 0.3593515 0.02414898
## 9  0.01192725      8 0.3168993 0.3470861 0.02398631
## 10 0.01192017      9 0.3049721 0.3410559 0.02393453
## 11 0.01182571     10 0.2930519 0.3410559 0.02393453
## 12 0.01000000     11 0.2812262 0.3328605 0.02381751
##
## Variable importance
##              movie05              movie10   exclusive_use_area
##                   15                   15                   12
##             office10             office15                 dong
##                   12                    8                    7
##             office05              movie15           hospital15
##                    6                    4                    3
##                kid15           hospital05 year_of_completion_f
##                    3                    3                    2
##                kid05             school15           hospital10
##                    2                    2                    2
##          bigMarket05                kid10                floor
##                    2                    1                    1
##
## Node number 1: 1455 observations,    complexity param=0.3135485
##   mean=1004.898, MSE=117570.8
##   left son=2 (1358 obs) right son=3 (97 obs)
##   Primary splits:
##       movie05           < 18.5     to the left,  improve=0.3135485, (0 missing)
##       movie10           < 55       to the left,  improve=0.3077123, (0 missing)
##       exclusive_use_area < 208.434  to the left,  improve=0.2776117, (0 missing)
##       dong              splits as  LLLLLLRRLLLRLLLRLLLLRRLL, improve=0.2229383, (0 missing)
##       hospital05        < 31.5     to the left,  improve=0.2212182, (0 missing)
##   Surrogate splits:
##       movie10           < 55       to the left,  agree=0.999, adj=0.990, (0 split)
##       office10          < 38.5     to the left,  agree=0.976, adj=0.639, (0 split)
##       exclusive_use_area < 208.434  to the left,  agree=0.973, adj=0.588, (0 split)
##       office05          < 13.5     to the left,  agree=0.959, adj=0.392, (0 split)
##       office15          < 69.5     to the left,  agree=0.945, adj=0.175, (0 split)
##
## Node number 2: 1358 observations,    complexity param=0.1410987
##   mean=953.5836, MSE=68859.17
##   left son=4 (967 obs) right son=5 (391 obs)
##   Primary splits:
##       office15   < 9.5      to the right, improve=0.2581212, (0 missing)
##       dong       splits as  LLLLLLRRLLLRLLLRLLLLRLLL, improve=0.2580890, (0 missing)
```

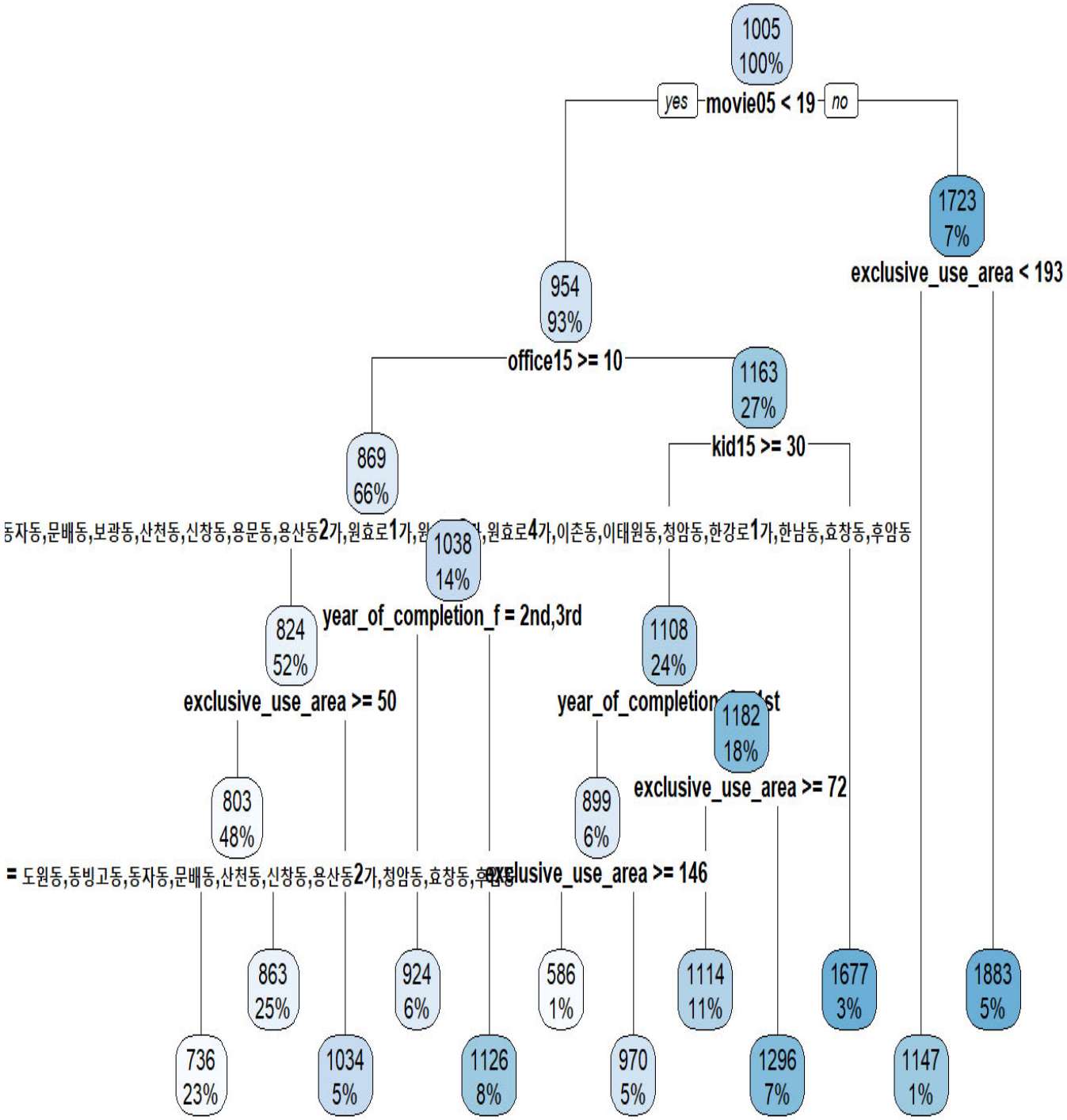
```
##       school15   < 7.5      to the right, improve=0.2335173, (0 missing)
##       office10   < 7.5      to the right, improve=0.2196412, (0 missing)
##       hospital15 < 136.5    to the right, improve=0.1846956, (0 missing)
##   Surrogate splits:
##       dong       splits as  LLLLLLLLLLLLLLLRLLLLLLLL, agree=0.912, adj=0.693, (0 split)
##       office10   < 5.5      to the right, agree=0.871, adj=0.552, (0 split)
##       school15   < 6.5      to the right, agree=0.798, adj=0.299, (0 split)
##       hospital15 < 106.5    to the right, agree=0.789, adj=0.266, (0 split)
##       movie15    < 28.5     to the right, agree=0.782, adj=0.243, (0 split)
##
## Node number 3: 97 observations,    complexity param=0.05211072
##   mean=1723.297, MSE=246571.3
##   left son=6 (21 obs) right son=7 (76 obs)
##   Primary splits:
##       exclusive_use_area < 193.121  to the left,  improve=0.3727136, (0 missing)
##       transaction_month  splits as  RLLLRRLRRRR,  improve=0.2394761, (0 missing)
##       floor              < 3.5      to the right, improve=0.0729411, (0 missing)
##   Surrogate splits:
##       bigMarket05 < 1        to the left,  agree=0.928, adj=0.667, (0 split)
##       hospital05  < 26.5     to the left,  agree=0.928, adj=0.667, (0 split)
##       movie05     < 22.5     to the right, agree=0.928, adj=0.667, (0 split)
##       movie10     < 58       to the right, agree=0.928, adj=0.667, (0 split)
##       movie15     < 79       to the right, agree=0.928, adj=0.667, (0 split)
##
## Node number 4: 967 observations,    complexity param=0.04321702
##   mean=868.8085, MSE=38944.5
##   left son=8 (763 obs) right son=9 (204 obs)
##   Primary splits:
##       dong                splits as  LLLLLLRRLLLRLLLLLLLRRLLL, improve=0.1963110, (0 missin
g)
##       year_of_completion_f splits as  LLLR, improve=0.1357723, (0 missing)
##       kid15               < 72.5     to the right, improve=0.1174880, (0 missing)
##       kid10               < 38       to the right, improve=0.1174880, (0 missing)
##       hospital15          < 190.5    to the right, improve=0.1162799, (0 missing)
##   Surrogate splits:
##       hospital05  < 41       to the left,  agree=0.846, adj=0.270, (0 split)
##       hospital15  < 41.5     to the right, agree=0.834, adj=0.211, (0 split)
##       hospital10  < 19.5     to the right, agree=0.824, adj=0.167, (0 split)
##       bigMarket05 < 2.5      to the left,  agree=0.814, adj=0.118, (0 split)
##       kid05       < 3.5      to the right, agree=0.800, adj=0.054, (0 split)
##
## Node number 5: 391 observations,    complexity param=0.06506483
##   mean=1163.245, MSE=81110.78
##   left son=10 (353 obs) right son=11 (38 obs)
##   Primary splits:
##       kid15               < 29.5     to the right, improve=0.3509566, (0 missing)
##       hospital15          < 55       to the right, improve=0.2021443, (0 missing)
##       movie15             < 15.5     to the right, improve=0.2021443, (0 missing)
##       year_of_completion_f splits as  LLLR,         improve=0.1855316, (0 missing)
##       floor               < 21.5     to the left,  improve=0.1099622, (0 missing)
##   Surrogate splits:
##       hospital15          < 55       to the right, agree=0.951, adj=0.500, (0 split)
##       movie15             < 15.5     to the right, agree=0.951, adj=0.500, (0 split)
##       kid05               < 15       to the left,  agree=0.939, adj=0.368, (0 split)
##       floor               < 23       to the left,  agree=0.923, adj=0.211, (0 split)
##       year_of_completion_f splits as  LLLR,         agree=0.923, adj=0.211, (0 split)
```

```
##
## Node number 6: 21 observations
##   mean=1146.589, MSE=266751.3
##
## Node number 7: 76 observations
##   mean=1882.651, MSE=123701.2
##
## Node number 8: 763 observations,    complexity param=0.01956231
##   mean=823.5971, MSE=31218.49
##   left son=16 (694 obs) right son=17 (69 obs)
##   Primary splits:
##       exclusive_use_area < 50.20805 to the right, improve=0.14049030, (0 missing)
##       dong               splits as  LLLLRL--LRL-RLRRRLR--RLL, improve=0.13303180, (0 missin
g)
##       school15           < 10.5     to the right, improve=0.09950408, (0 missing)
##       subway10           < 2.5      to the right, improve=0.09863773, (0 missing)
##       hospital15         < 160      to the right, improve=0.09124124, (0 missing)
##   Surrogate splits:
##       office15 < 77       to the left,  agree=0.924, adj=0.159, (0 split)
##       movie10  < 47.5     to the left,  agree=0.912, adj=0.029, (0 split)
##       dong     splits as  LLLLLL--LLL-LRLLLLL--LLL, agree=0.911, adj=0.014, (0 split)
##
## Node number 9: 204 observations,    complexity param=0.01192017
##   mean=1037.908, MSE=31601.38
##   left son=18 (89 obs) right son=19 (115 obs)
##   Primary splits:
##       year_of_completion_f splits as  RLLR,         improve=0.3163069, (0 missing)
##       school05             < 0.5      to the left,  improve=0.1704078, (0 missing)
##       hospital10           < 83.5     to the left,  improve=0.1587010, (0 missing)
##       bigMarket05          < 2.5      to the right, improve=0.1354760, (0 missing)
##       movie05              < 8.5      to the right, improve=0.1354760, (0 missing)
##   Surrogate splits:
##       dong       splits as  ------RR---R-------LL---, agree=0.897, adj=0.764, (0 split)
##       hospital10 < 89.5     to the left,  agree=0.853, adj=0.663, (0 split)
##       movie10    < 19       to the left,  agree=0.848, adj=0.652, (0 split)
##       kid10      < 24.5     to the left,  agree=0.833, adj=0.618, (0 split)
##       movie05    < 3.5      to the right, agree=0.819, adj=0.584, (0 split)
##
## Node number 10: 353 observations,    complexity param=0.03205798
##   mean=1107.888, MSE=49474.38
##   left son=20 (93 obs) right son=21 (260 obs)
##   Primary splits:
##       year_of_completion_f splits as  LRRR,         improve=0.3140100, (0 missing)
##       kid10                < 24.5     to the left,  improve=0.2094978, (0 missing)
##       exclusive_use_area   < 67.905   to the right, improve=0.1811225, (0 missing)
##       hospital15           < 99.5     to the left,  improve=0.1439076, (0 missing)
##       office05             < 2.5      to the left,  improve=0.1422714, (0 missing)
##   Surrogate splits:
##       kid10      < 24.5     to the left,  agree=0.844, adj=0.409, (0 split)
##       hospital10 < 47       to the left,  agree=0.830, adj=0.355, (0 split)
##       hospital05 < 20.5     to the left,  agree=0.827, adj=0.344, (0 split)
##       kid05      < 9.5      to the left,  agree=0.827, adj=0.344, (0 split)
##       office05   < 1.5      to the left,  agree=0.822, adj=0.323, (0 split)
##
## Node number 11: 38 observations
##   mean=1677.479, MSE=82092.38
```

```
##
## Node number 16: 694 observations,    complexity param=0.01644067
##   mean=802.715, MSE=23505.81
##   left son=32 (330 obs) right son=33 (364 obs)
##   Primary splits:
##       dong       splits as  LLLLRL--LRL-R-RRRLR--RLL, improve=0.1724040, (0 missing)
##       kid05      < 7.5      to the right, improve=0.1173198, (0 missing)
##       hospital10 < 85.5     to the right, improve=0.1094987, (0 missing)
##       subway10   < 2.5      to the right, improve=0.1053347, (0 missing)
##       school15   < 10.5     to the right, improve=0.1023676, (0 missing)
##   Surrogate splits:
##       hospital10 < 85.5     to the right, agree=0.912, adj=0.815, (0 split)
##       kid05      < 10.5     to the right, agree=0.854, adj=0.694, (0 split)
##       kid15      < 47.5     to the right, agree=0.817, adj=0.615, (0 split)
##       school10   < 4.5      to the right, agree=0.816, adj=0.612, (0 split)
##       subway15   < 6.5      to the right, agree=0.801, adj=0.582, (0 split)
##
## Node number 17: 69 observations
##   mean=1033.629, MSE=60293.27
##
## Node number 18: 89 observations
##   mean=924.2603, MSE=18021.11
##
## Node number 19: 115 observations
##   mean=1125.862, MSE=24379.76
##
## Node number 20: 93 observations,    complexity param=0.01192725
##   mean=899.4833, MSE=52162.95
##   left son=40 (17 obs) right son=41 (76 obs)
##   Primary splits:
##       exclusive_use_area < 146.465  to the right, improve=0.4205887, (0 missing)
##       office15           < 7.5      to the left,  improve=0.1942246, (0 missing)
##       school05           < 2.5      to the right, improve=0.1787468, (0 missing)
##       transaction_month  splits as  LLLLLLRRRRR,  improve=0.1780058, (0 missing)
##       movie15            < 18.5     to the left,  improve=0.1305483, (0 missing)
##   Surrogate splits:
##       kid15      < 32       to the left,  agree=0.860, adj=0.235, (0 split)
##       hospital05 < 31.5     to the right, agree=0.849, adj=0.176, (0 split)
##       movie05    < 1.5      to the right, agree=0.849, adj=0.176, (0 split)
##       kid05      < 16       to the right, agree=0.849, adj=0.176, (0 split)
##       movie10    < 14.5     to the right, agree=0.839, adj=0.118, (0 split)
##
## Node number 21: 260 observations,    complexity param=0.01182571
##   mean=1182.433, MSE=27420.33
##   left son=42 (162 obs) right son=43 (98 obs)
##   Primary splits:
##       exclusive_use_area < 71.805   to the right, improve=0.2837550, (0 missing)
##       transaction_month  splits as  LLLLLRRLRRR,  improve=0.1949131, (0 missing)
##       subway15           < 5.5      to the left,  improve=0.1363305, (0 missing)
##       hospital05         < 32       to the left,  improve=0.1284009, (0 missing)
##       office05           < 2.5      to the left,  improve=0.1044675, (0 missing)
##   Surrogate splits:
##       bigMarket10 < 0.5      to the right, agree=0.673, adj=0.133, (0 split)
##       subway10    < 1.5      to the right, agree=0.673, adj=0.133, (0 split)
##       hospital10  < 24       to the right, agree=0.673, adj=0.133, (0 split)
##       movie10     < 6.5      to the right, agree=0.673, adj=0.133, (0 split)
```

```
##       kid05       < 5.5      to the right, agree=0.673, adj=0.133, (0 split)
##
## Node number 32: 330 observations
##   mean=735.8568, MSE=15005.75
##
## Node number 33: 364 observations
##   mean=863.3283, MSE=23485.45
##
## Node number 40: 17 observations
##   mean=586.3046, MSE=2602.078
##
## Node number 41: 76 observations
##   mean=969.5364, MSE=36402.35
##
## Node number 42: 162 observations
##   mean=1113.826, MSE=21594.65
##
## Node number 43: 98 observations
##   mean=1295.843, MSE=16407.99
```

```
# install.packages("rpart.plot")
library(rpart.plot)

rpart.plot(tree1, cex = 0.7)
```

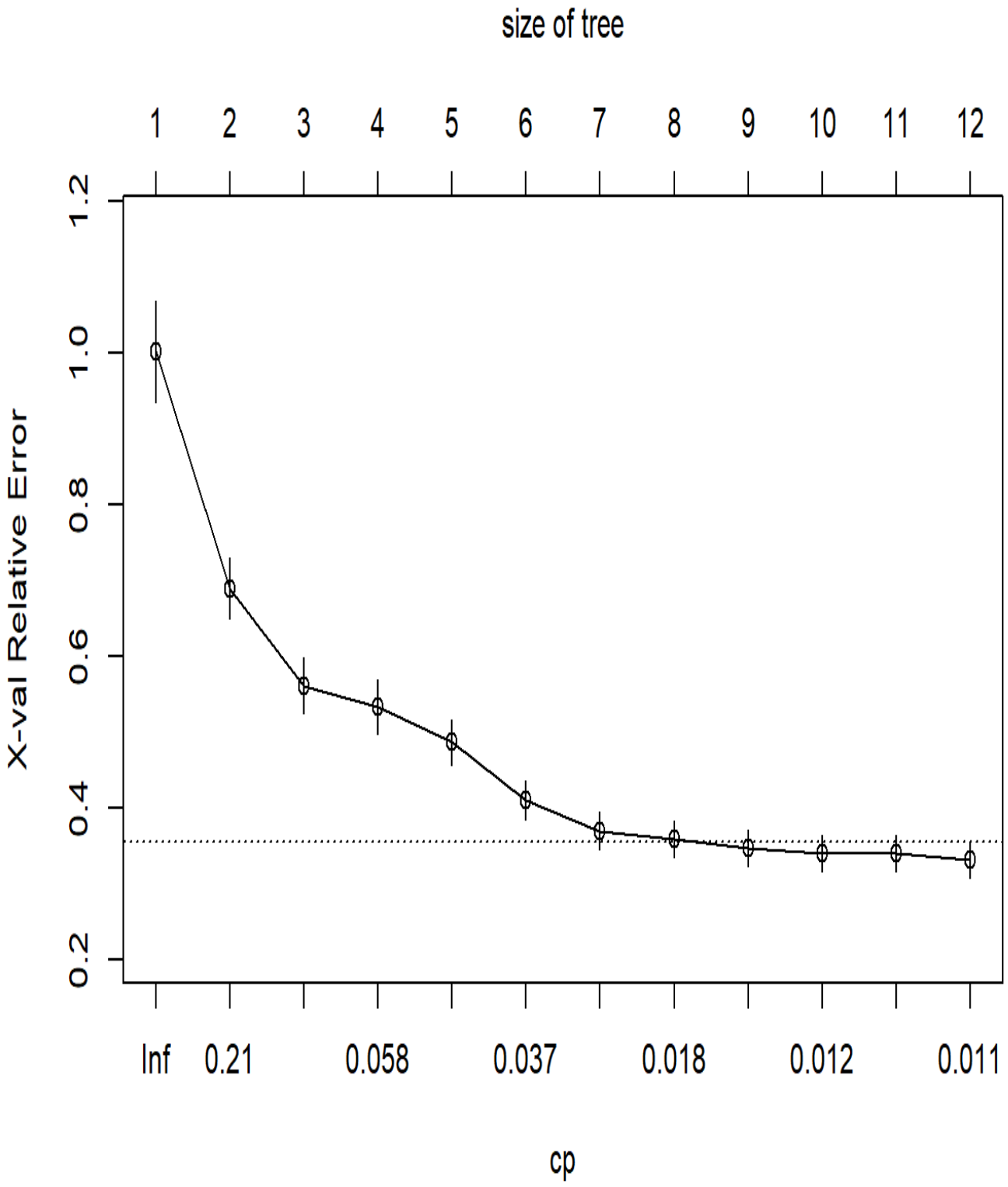


Decision Tree parameter tuning

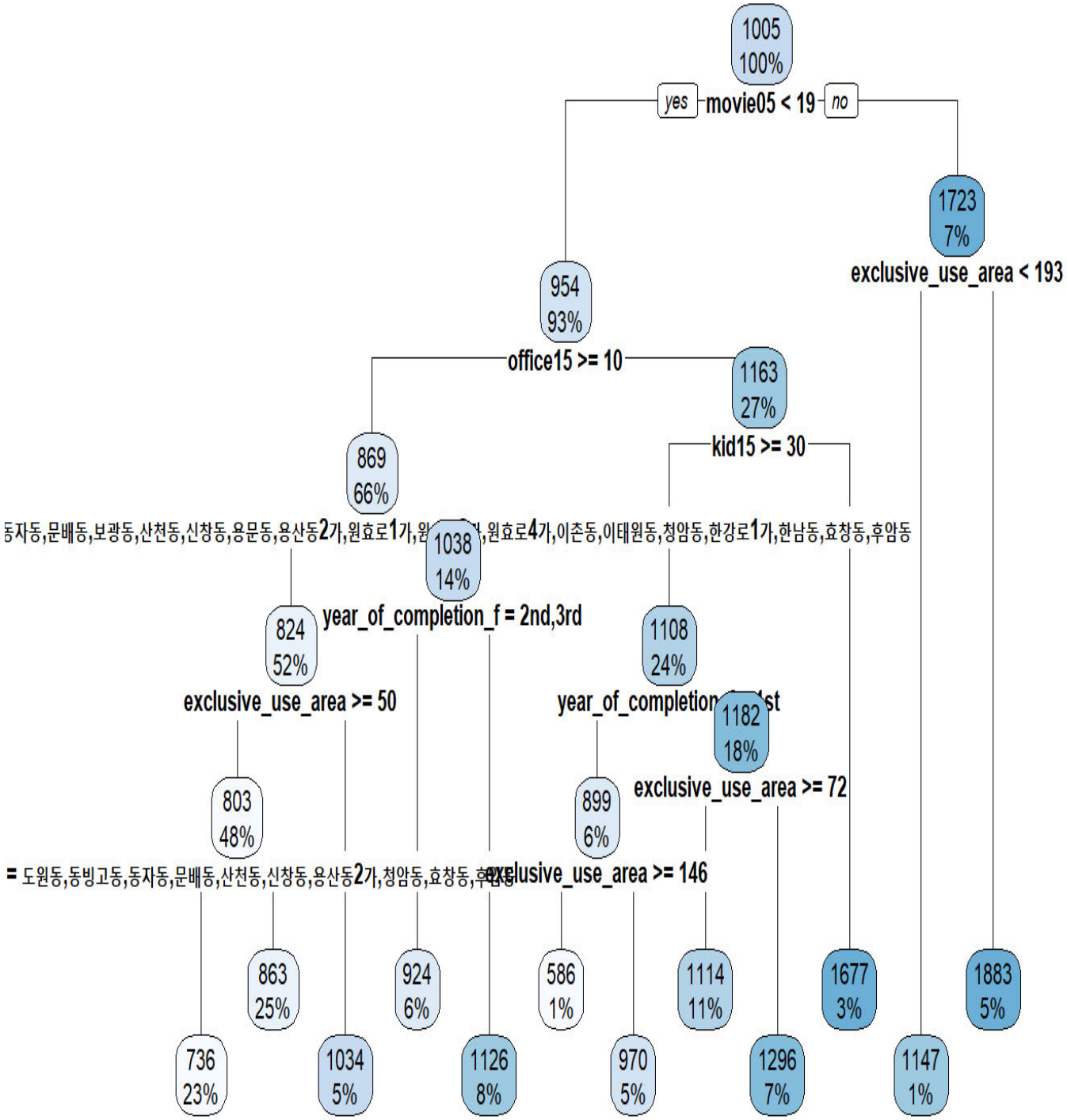
```
printcp(tree1)
```

```
##
## Regression tree:
## rpart(formula = unit_price ~ . - year, data = data_train1, method = "anova",
##     control = rpart.control(minsplit = 50, maxdepth = 5))
##
## Variables actually used in tree construction:
## [1] dong                 exclusive_use_area   kid15
## [4] movie05              office15             year_of_completion_f
##
## Root node error: 171065491/1455 = 117571
##
## n= 1455
##
##          CP nsplit rel error  xerror     xstd
## 1  0.313548      0   1.00000 1.00247 0.066323
## 2  0.141099      1   0.68645 0.68971 0.040205
## 3  0.065065      2   0.54535 0.56204 0.036761
## 4  0.052111      3   0.48029 0.53321 0.035824
## 5  0.043217      4   0.42818 0.48715 0.029294
## 6  0.032058      5   0.38496 0.41147 0.025583
## 7  0.019562      6   0.35290 0.37061 0.024297
## 8  0.016441      7   0.33334 0.35935 0.024149
## 9  0.011927      8   0.31690 0.34709 0.023986
## 10 0.011920      9   0.30497 0.34106 0.023935
## 11 0.011826     10   0.29305 0.34106 0.023935
## 12 0.010000     11   0.28123 0.33286 0.023818
```

```
plotcp(tree1)
```




```
tree1 <- prune(tree1, cp= tree1$cptable[which.min(tree1$cptable[, "xerror"]), "CP"])

rpart.plot(tree1, cex = 0.7)
```



Decision Tree prediction & RMSE calculation

```
# test data 에 적용

predict_1 <- predict(tree1, data_test1)
summary(predict_1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   586.3   863.3   924.3  1002.3  1113.8  1882.7
```

```
# actual, predicted cbind

databind1 <- cbind(data_test1[,25],predict_1)
databind1 <- as.data.frame(databind1)
summary(databind1)
```

```
##        V1           predict_1
##  Min.   : 362.5   Min.   : 586.3
##  1st Qu.: 779.3   1st Qu.: 863.3
##  Median : 937.4   Median : 924.3
##  Mean   :1016.1   Mean   :1002.3
##  3rd Qu.:1154.1   3rd Qu.:1113.8
##  Max.   :2983.9   Max.   :1882.7
```

```
# RMSE 계산
install.packages("Metrics", repos ="http://cran.us.r-project.org")
```

```
## 패키지 'Metrics'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:₩Users₩LUIS₩AppData₩Local₩Temp₩Rtmpkp5M2w₩downloaded_packages
```

```
library(Metrics)
rmse(databind1$V1, databind1$predict_1)
```

```
## [1] 209.9278
```

Linear regression

```
# factor 변수 중 unique value 있는지 찾아보기
str(data_train1)
```

```
## 'data.frame':    1455 obs. of  28 variables:
##  $ dong                : Factor w/ 24 levels "도원동","동빙고동",..: 24 13 6 18 23 23 23 23
1 10 ...
##  $ exclusive_use_area  : num  62.3 46 85 223.8 59.4 ...
##  $ floor               : int  6 3 16 9 17 9 9 6 1 6 ...
##  $ bigMarket05         : int  0 1 0 0 1 1 1 0 0 0 ...
##  $ bigMarket10         : int  2 2 2 3 4 4 4 4 5 5 ...
##  $ bigMarket15         : int  2 4 6 5 8 8 8 8 8 7 ...
##  $ school05            : int  0 5 4 2 3 3 3 5 3 2 ...
##  $ school10            : int  7 14 5 5 12 12 13 14 11 10 ...
##  $ school15            : int  28 21 12 13 26 26 28 29 24 19 ...
##  $ subway05            : int  1 1 0 0 3 3 3 2 2 2 ...
##  $ subway10            : int  6 6 3 3 6 6 6 8 6 7 ...
##  $ subway15            : int  12 11 9 8 13 13 13 15 14 14 ...
##  $ hospital05          : int  15 39 13 14 20 20 35 14 24 31 ...
##  $ hospital10          : int  89 81 117 107 182 182 189 132 169 98 ...
##  $ hospital15          : int  266 158 203 202 304 304 296 299 316 293 ...
##  $ movie05             : int  8 4 1 2 6 6 6 8 5 6 ...
##  $ movie10             : int  27 27 9 6 18 18 18 18 17 17 ...
##  $ movie15             : int  89 49 21 17 32 32 35 42 36 35 ...
##  $ kid05               : int  11 6 16 13 21 21 22 13 18 18 ...
##  $ kid10               : int  29 27 41 40 66 66 66 59 57 53 ...
##  $ kid15               : int  57 70 81 73 113 113 115 102 100 93 ...
##  $ office05            : int  8 5 1 1 3 3 3 3 2 1 ...
##  $ office10            : int  22 13 9 10 15 15 15 13 13 12 ...
##  $ office15            : int  68 21 18 16 32 32 31 33 32 27 ...
##  $ unit_price          : num  787 1152 671 773 1052 ...
##  $ transaction_month   : Factor w/ 11 levels "01","02","03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year                : Factor w/ 42 levels "1969","1970",..: 11 39 26 30 35 35 31 1 3 36
...
##  $ year_of_completion_f: Factor w/ 4 levels "1st","2nd","3rd",..: 1 4 2 3 4 4 3 1 1 4 ...
```

```
sapply(lapply(data_train1, unique), length)
```

```
##                 dong   exclusive_use_area                floor
##                   24                  373                   43
##          bigMarket05         bigMarket10         bigMarket15
##                    4                    6                   10
##             school05            school10            school15
##                    7                   17                   27
##             subway05            subway10            subway15
##                    4                   10                   15
##           hospital05          hospital10          hospital15
##                   42                   84                  100
##              movie05             movie10             movie15
##                   26                   46                   63
##                kid05               kid10               kid15
##                   21                   45                   61
##             office05            office10            office15
##                   20                   42                   51
##           unit_price    transaction_month                year
##                 1257                   11                   42
## year_of_completion_f
##                    4
```

```
# Linear Model  (dong은 제외하고 분석:삭제)
linear1 <- lm(unit_price ~.-year, data = data_train1)
#linear1 <- lm(unit_price ~ dong+exclusive_use_area+floor+bigMarket05+bigMarket10+bigMarket15+school05+school10+school15+subway05+subway10+subway15+hospital05+hospital10+hospital15+movie05+movie10+movie15+kid05+kid10+kid15+office05+office10+office15+transaction_month+year_of_completion_f, data = data_train1)

summary(linear1)
```

```
##
## Call:
## lm(formula = unit_price ~ . - year, data = data_train1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -801.82  -98.84    3.26   99.08 1491.48
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              539.2301   267.8252   2.013 0.044268 *
## dong동빙고동             313.3528   241.4995   1.298 0.194664
## dong동자동               465.7723   202.8373   2.296 0.021807 *
## dong문배동                22.1725   162.9295   0.136 0.891772
## dong보광동               514.4286   218.3628   2.356 0.018618 *
## dong산천동              -379.8959   106.7629  -3.558 0.000386 ***
## dong서빙고동             651.1875   216.2973   3.011 0.002654 **
## dong신계동               -70.2135   167.2656  -0.420 0.674716
## dong신창동              -349.3010   110.7719  -3.153 0.001649 **
## dong용문동              -699.4860   156.8824  -4.459 8.91e-06 ***
## dong용산동2가            227.9760   250.0240   0.912 0.362023
## dong용산동5가           -254.5126   194.8450  -1.306 0.191689
## dong원효로1가           -385.6282   165.0308  -2.337 0.019595 *
## dong원효로2가           -132.9360   263.1068  -0.505 0.613460
## dong원효로4가           -104.3265   134.4093  -0.776 0.437771
## dong이촌동               -10.5846   195.4177  -0.054 0.956812
## dong이태원동             534.6793   241.3089   2.216 0.026870 *
## dong청암동              -275.7889   164.9723  -1.672 0.094802 .
## dong한강로1가            139.8990   193.4260   0.723 0.469636
## dong한강로2가            -77.4476   221.8490  -0.349 0.727067
## dong한강로3가           -271.1485   193.1787  -1.404 0.160656
## dong한남동               379.5173   238.2639   1.593 0.111422
## dong효창동              -205.5584    88.3107  -2.328 0.020072 *
## dong후암동               529.4795   194.1190   2.728 0.006460 **
## exclusive_use_area        -0.1907     0.1574  -1.212 0.225755
## floor                      4.2348     0.8489   4.989 6.84e-07 ***
## bigMarket05             -142.6870    24.6489  -5.789 8.74e-09 ***
## bigMarket10               -5.9804    15.8257  -0.378 0.705569
## bigMarket15              104.0315    13.8427   7.515 1.01e-13 ***
## school05                 -13.6817    13.8171  -0.990 0.322248
## school10                 -13.9323     8.0670  -1.727 0.084374 .
## school15                  -7.3374     7.5751  -0.969 0.332901
## subway05                 -65.2358    14.6218  -4.462 8.79e-06 ***
## subway10                 -37.9352    11.2068  -3.385 0.000731 ***
## subway15                  64.0398    10.6383   6.020 2.23e-09 ***
## hospital05                12.5939     1.6543   7.613 4.92e-14 ***
## hospital10                 0.8244     0.9662   0.853 0.393653
## hospital15                -4.3926     0.8413  -5.221 2.05e-07 ***
## movie05                   13.2979     2.9816   4.460 8.85e-06 ***
## movie10                   -2.0011     2.4406  -0.820 0.412393
## movie15                   -1.4007     2.2767  -0.615 0.538489
## kid05                     18.8887     4.7976   3.937 8.65e-05 ***
## kid10                      2.5359     2.8306   0.896 0.370466
## kid15                     -0.4428     2.3103  -0.192 0.848036
## office05                   4.0935     4.6872   0.873 0.382629
```

```
## office10                -12.0083     3.4623  -3.468 0.000540 ***
## office15                 -0.9249     2.8986  -0.319 0.749721
## transaction_month02      29.9018    36.0235   0.830 0.406644
## transaction_month03      48.4084    34.7057   1.395 0.163291
## transaction_month04      28.6261    34.3123   0.834 0.404265
## transaction_month05      71.6628    31.8804   2.248 0.024741 *
## transaction_month06      90.2086    33.2640   2.712 0.006772 **
## transaction_month07     112.8317    33.1058   3.408 0.000673 ***
## transaction_month08     154.3360    40.6155   3.800 0.000151 ***
## transaction_month09     174.3746    37.3439   4.669 3.31e-06 ***
## transaction_month10     152.4918    38.3433   3.977 7.34e-05 ***
## transaction_month11     242.2804    36.7464   6.593 6.09e-11 ***
## year_of_completion_f2nd   8.2579    21.2459   0.389 0.697569
## year_of_completion_f3rd 141.4504    26.9701   5.245 1.81e-07 ***
## year_of_completion_f4th 578.2144    27.3910  21.110  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 199.9 on 1395 degrees of freedom
## Multiple R-squared:  0.674,  Adjusted R-squared:  0.6602
## F-statistic: 48.88 on 59 and 1395 DF,  p-value: < 2.2e-16
```

```
print(linear1)
```

```
##
## Call:
## lm(formula = unit_price ~ . - year, data = data_train1)
##
## Coefficients:
##             (Intercept)            dong동빙고동               dong동자동
##                539.2301                313.3528                465.7723
##               dong문배동               dong보광동               dong산천동
##                 22.1725                514.4286               -379.8959
##            dong서빙고동               dong신계동               dong신창동
##                651.1875                -70.2135               -349.3010
##               dong용문동            dong용산동2가            dong용산동5가
##               -699.4860                227.9760               -254.5126
##            dong원효로1가            dong원효로2가            dong원효로4가
##               -385.6282               -132.9360               -104.3265
##               dong이촌동              dong이태원동               dong청암동
##                -10.5846                534.6793               -275.7889
##            dong한강로1가            dong한강로2가            dong한강로3가
##                139.8990                -77.4476               -271.1485
##               dong한남동               dong효창동               dong후암동
##                379.5173               -205.5584                529.4795
##      exclusive_use_area                   floor              bigMarket05
##                 -0.1907                  4.2348                -142.6870
##             bigMarket10              bigMarket15                 school05
##                 -5.9804                104.0315                 -13.6817
##                school10                school15                 subway05
##                -13.9323                 -7.3374                 -65.2358
##                subway10                subway15                hospital05
##                -37.9352                 64.0398                  12.5939
##             hospital10              hospital15                   movie05
##                  0.8244                 -4.3926                  13.2979
##                movie10                 movie15                    kid05
##                 -2.0011                 -1.4007                  18.8887
##                  kid10                   kid15                office05
##                  2.5359                 -0.4428                   4.0935
##                office10                office15      transaction_month02
##                -12.0083                 -0.9249                  29.9018
##     transaction_month03      transaction_month04      transaction_month05
##                 48.4084                 28.6261                  71.6628
##     transaction_month06      transaction_month07      transaction_month08
##                 90.2086                112.8317                 154.3360
##     transaction_month09      transaction_month10      transaction_month11
##                174.3746                152.4918                 242.2804
## year_of_completion_f2nd  year_of_completion_f3rd  year_of_completion_f4th
##                  8.2579                141.4504                 578.2144
```

```
linear1$coefficients
```

```
##             (Intercept)            dong동빙고동             dong동자동
##               539.2300700               313.3527543               465.7722608
##              dong문배동              dong보광동              dong산천동
##                22.1725266               514.4286135              -379.8959471
##             dong서빙고동              dong신계동              dong신창동
##               651.1874808               -70.2135424              -349.3010060
##              dong용문동            dong용산동2가            dong용산동5가
##              -699.4859765               227.9760297              -254.5126290
##            dong원효로1가            dong원효로2가            dong원효로4가
##              -385.6282170              -132.9360035              -104.3265422
##              dong이촌동            dong이태원동              dong청암동
##               -10.5846212               534.6793313              -275.7889465
##            dong한강로1가            dong한강로2가            dong한강로3가
##               139.8989710               -77.4475919              -271.1484937
##              dong한남동              dong효창동              dong후암동
##               379.5173167              -205.5583832               529.4794973
##      exclusive_use_area                     floor               bigMarket05
##                -0.1907333                 4.2347623              -142.6870325
##              bigMarket10               bigMarket15                  school05
##                -5.9804104               104.0314765               -13.6817154
##                 school10                  school15                  subway05
##               -13.9323046                -7.3373983               -65.2357987
##                 subway10                  subway15                 hospital05
##               -37.9351680                64.0398309                12.5939047
##                hospital10                hospital15                   movie05
##                 0.8244488                -4.3926342                13.2979314
##                   movie10                   movie15                     kid05
##                -2.0011443                -1.4007459                18.8886949
##                    kid10                     kid15                  office05
##                 2.5358830                -0.4427998                 4.0934695
##                  office10                  office15     transaction_month02
##               -12.0082816                -0.9248652                29.9018413
##     transaction_month03     transaction_month04     transaction_month05
##                48.4083524                28.6261276                71.6627640
##     transaction_month06     transaction_month07     transaction_month08
##                90.2085769               112.8317273               154.3359646
##     transaction_month09     transaction_month10     transaction_month11
##               174.3746424               152.4918108               242.2804276
## year_of_completion_f2nd year_of_completion_f3rd year_of_completion_f4th
##                 8.2579239               141.4503846               578.2144045
```

Linear regression parameter tuning

```
step(linear1, direction = "both")
```

```
## Start:  AIC=15476
## unit_price ~ (dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket10 + bigMarket15 + school05 + school10 + school15 +
##     subway05 + subway10 + subway15 + hospital05 + hospital10 +
##     hospital15 + movie05 + movie10 + movie15 + kid05 + kid10 +
##     kid15 + office05 + office10 + office15 + transaction_month +
##     year + year_of_completion_f) - year
##
##                        Df Sum of Sq      RSS   AIC
## - kid15                 1      1469 55769667 15474
## - office15              1      4070 55772268 15474
## - bigMarket10           1      5709 55773907 15474
## - movie15               1     15133 55783331 15474
## - movie10               1     26876 55795075 15475
## - hospital10            1     29107 55797305 15475
## - office05              1     30491 55798690 15475
## - kid10                 1     32086 55800285 15475
## - school15              1     37508 55805706 15475
## - school05              1     39198 55807396 15475
## - exclusive_use_area    1     58715 55826913 15476
## <none>                              55768198 15476
## - school10              1    119245 55887443 15477
## - subway10              1    458075 56226273 15486
## - office10              1    480890 56249088 15486
## - kid05                 1    619677 56387876 15490
## - movie05               1    795224 56563422 15495
## - subway05              1    795764 56563962 15495
## - floor                 1    994936 56763135 15500
## - hospital15            1   1089801 56857999 15502
## - bigMarket05           1   1339639 57107837 15508
## - subway15              1   1448675 57216873 15511
## - bigMarket15           1   2257874 58026073 15532
## - hospital05            1   2316777 58084975 15533
## - transaction_month    10   4728730 60496928 15574
## - dong                 23  11548301 67316499 15704
## - year_of_completion_f  3  20903787 76671985 15933
##
## Step:  AIC=15474.04
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket10 + bigMarket15 + school05 + school10 + school15 +
##     subway05 + subway10 + subway15 + hospital05 + hospital10 +
##     hospital15 + movie05 + movie10 + movie15 + kid05 + kid10 +
##     office05 + office10 + office15 + transaction_month + year_of_completion_f
##
##                        Df Sum of Sq      RSS   AIC
## - office15              1      3416 55773083 15472
## - bigMarket10           1      4943 55774610 15472
## - movie15               1     15055 55784722 15472
## - movie10               1     25758 55795425 15473
## - hospital10            1     29301 55798968 15473
## - office05              1     30053 55799720 15473
## - kid10                 1     30989 55800656 15473
## - school05              1     38290 55807957 15473
## - school15              1     46601 55816267 15473
## - exclusive_use_area    1     60239 55829905 15474
```

```
## <none>                              55769667 15474
## - school10              1    126667 55896333 15475
## + kid15                 1      1469 55768198 15476
## - subway10              1    491196 56260863 15485
## - office10              1    512704 56282371 15485
## - kid05                 1    662031 56431698 15489
## - movie05               1    796703 56566370 15493
## - subway05              1    842526 56612193 15494
## - floor                 1    997339 56767006 15498
## - hospital15            1   1184072 56953738 15503
## - subway15              1   1447207 57216874 15509
## - bigMarket05           1   1615209 57384876 15514
## - hospital05            1   2345623 58115290 15532
## - bigMarket15           1   2366522 58136189 15532
## - transaction_month    10   4727564 60497231 15572
## - dong                 23  13905575 69675242 15752
## - year_of_completion_f  3  20905137 76674803 15931
##
## Step:  AIC=15472.13
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket10 + bigMarket15 + school05 + school10 + school15 +
##     subway05 + subway10 + subway15 + hospital05 + hospital10 +
##     hospital15 + movie05 + movie10 + movie15 + kid05 + kid10 +
##     office05 + office10 + transaction_month + year_of_completion_f
##
##                        Df Sum of Sq      RSS   AIC
## - bigMarket10           1      4618 55777701 15470
## - movie15               1     16890 55789973 15471
## - kid10                 1     29208 55802291 15471
## - office05              1     29985 55803068 15471
## - movie10               1     31672 55804755 15471
## - school05              1     35494 55808577 15471
## - hospital10            1     35942 55809025 15471
## - school15              1     46997 55820080 15471
## - exclusive_use_area    1     64735 55837818 15472
## <none>                              55773083 15472
## - school10              1    123396 55896479 15473
## + office15              1      3416 55769667 15474
## + kid15                 1       815 55772268 15474
## - subway10              1    502474 56275556 15483
## - kid05                 1    699611 56472694 15488
## - office10              1    765644 56538727 15490
## - subway05              1    842779 56615861 15492
## - floor                 1    994259 56767341 15496
## - movie05               1   1130531 56903614 15499
## - hospital15            1   1220616 56993699 15502
## - subway15              1   1443795 57216878 15507
## - bigMarket05           1   1633202 57406284 15512
## - hospital05            1   2408187 58181270 15532
## - bigMarket15           1   2639044 58412127 15537
## - transaction_month    10   4744624 60517707 15571
## - dong                 23  13904901 69677983 15750
## - year_of_completion_f  3  20972033 76745116 15931
##
## Step:  AIC=15470.25
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
```

```
##     bigMarket15 + school05 + school10 + school15 + subway05 +
##     subway10 + subway15 + hospital05 + hospital10 + hospital15 +
##     movie05 + movie10 + movie15 + kid05 + kid10 + office05 +
##     office10 + transaction_month + year_of_completion_f
##
##                         Df Sum of Sq      RSS   AIC
## - movie15                1     16790 55794491 15469
## - office05               1     25968 55803669 15469
## - kid10                  1     30103 55807803 15469
## - movie10                1     30996 55808697 15469
## - hospital10             1     31343 55809044 15469
## - school05               1     41851 55819552 15469
## - school15               1     44880 55822580 15469
## - exclusive_use_area     1     66626 55844326 15470
## <none>                               55777701 15470
## - school10               1    118780 55896481 15471
## + bigMarket10            1      4618 55773083 15472
## + office15               1      3091 55774610 15472
## + kid15                  1       318 55777383 15472
## - subway10               1    670518 56448219 15486
## - kid05                  1    705406 56483107 15486
## - office10               1    779310 56557010 15488
## - subway05               1    843228 56620929 15490
## - floor                  1   1015362 56793063 15494
## - movie05                1   1128782 56906482 15497
## - hospital15             1   1322738 57100438 15502
## - subway15               1   1439350 57217051 15505
## - bigMarket05            1   1635185 57412885 15510
## - hospital05             1   2404329 58182030 15530
## - bigMarket15            1   2660171 58437872 15536
## - transaction_month     10   4740165 60517865 15569
## - dong                  23  13903292 69680993 15748
## - year_of_completion_f   3  21044594 76822295 15930
##
## Step:  AIC=15468.69
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket15 + school05 + school10 + school15 + subway05 +
##     subway10 + subway15 + hospital05 + hospital10 + hospital15 +
##     movie05 + movie10 + kid05 + kid10 + office05 + office10 +
##     transaction_month + year_of_completion_f
##
##                         Df Sum of Sq      RSS   AIC
## - hospital10             1     22605 55817096 15467
## - movie10                1     32765 55827256 15468
## - office05               1     42979 55837470 15468
## - school15               1     57506 55851997 15468
## - school05               1     65650 55860141 15468
## - exclusive_use_area     1     69132 55863623 15468
## - kid10                  1     72425 55866916 15469
## <none>                               55794491 15469
## - school10               1    126821 55921312 15470
## + movie15                1     16790 55777701 15470
## + office15               1      4845 55789646 15471
## + bigMarket10            1      4519 55789973 15471
## + kid15                  1       213 55794278 15471
## - subway10               1    737325 56531816 15486
```

```
## - kid05                 1    773187 56567679 15487
## - subway05              1    866732 56661223 15489
## - office10              1   1004226 56798717 15493
## - floor                 1   1012312 56806803 15493
## - movie05               1   1118000 56912491 15496
## - subway15              1   1429268 57223759 15504
## - hospital15            1   1432951 57227442 15504
## - bigMarket05           1   1621278 57415770 15508
## - hospital05            1   2455177 58249668 15529
## - bigMarket15           1   2679732 58474223 15535
## - transaction_month    10   4743010 60537501 15567
## - dong                 23  14376781 70171272 15756
## - year_of_completion_f  3  21119844 76914335 15930
##
## Step:  AIC=15467.28
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket15 + school05 + school10 + school15 + subway05 +
##     subway10 + subway15 + hospital05 + hospital15 + movie05 +
##     movie10 + kid05 + kid10 + office05 + office10 + transaction_month +
##     year_of_completion_f
##
##                        Df Sum of Sq      RSS   AIC
## - office05              1     27932 55845027 15466
## - movie10               1     47659 55864755 15466
## - exclusive_use_area    1     70457 55887553 15467
## <none>                              55817096 15467
## - school05              1     79642 55896737 15467
## - school15              1     88578 55905674 15468
## - kid10                 1    112786 55929882 15468
## - school10              1    119672 55936768 15468
## + hospital10            1     22605 55794491 15469
## + office15              1     10812 55806284 15469
## + movie15               1      8052 55809044 15469
## + kid15                 1       460 55816636 15469
## + bigMarket10           1        13 55817083 15469
## - subway10              1    742599 56559695 15484
## - kid05                 1    844398 56661494 15487
## - subway05              1    852443 56669539 15487
## - floor                 1   1016345 56833441 15492
## - movie05               1   1100740 56917836 15494
## - office10              1   1137272 56954368 15495
## - hospital15            1   1429643 57246739 15502
## - subway15              1   1566594 57383690 15506
## - bigMarket05           1   1749425 57566520 15510
## - hospital05            1   2474687 58291783 15528
## - bigMarket15           1   2661826 58478922 15533
## - transaction_month    10   4733829 60550925 15566
## - dong                 23  14526219 70343315 15758
## - year_of_completion_f  3  21121490 76938586 15928
##
## Step:  AIC=15466
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket15 + school05 + school10 + school15 + subway05 +
##     subway10 + subway15 + hospital05 + hospital15 + movie05 +
##     movie10 + kid05 + kid10 + office10 + transaction_month +
##     year_of_completion_f
```

```
##
##                         Df Sum of Sq      RSS   AIC
## - movie10                1     33984 55879012 15465
## - exclusive_use_area     1     68641 55913669 15466
## <none>                               55845027 15466
## - school15               1    108586 55953613 15467
## - school05               1    122193 55967220 15467
## + office05               1     27932 55817096 15467
## + movie15                1     20930 55824097 15468
## - school10               1    143055 55988082 15468
## + office15               1      9372 55835656 15468
## + hospital10             1      7557 55837470 15468
## - kid10                  1    149897 55994924 15468
## + bigMarket10            1       359 55844668 15468
## + kid15                  1       337 55844691 15468
## - kid05                  1    818384 56663412 15485
## - subway10               1    841399 56686426 15486
## - subway05               1    893854 56738881 15487
## - floor                  1   1027910 56872938 15490
## - movie05                1   1164315 57009343 15494
## - office10               1   1204388 57049415 15495
## - hospital15             1   1417692 57262720 15500
## - subway15               1   1624702 57469730 15506
## - bigMarket05            1   1744072 57589100 15509
## - bigMarket15            1   2685969 58530997 15532
## - hospital05             1   2823469 58668497 15536
## - transaction_month     10   4728028 60573055 15564
## - dong                  23  14943374 70788402 15765
## - year_of_completion_f   3  21175122 77020149 15928
##
## Step:  AIC=15464.89
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket15 + school05 + school10 + school15 + subway05 +
##     subway10 + subway15 + hospital05 + hospital15 + movie05 +
##     kid05 + kid10 + office10 + transaction_month + year_of_completion_f
##
##                         Df Sum of Sq      RSS   AIC
## - exclusive_use_area     1     71720 55950732 15465
## <none>                               55879012 15465
## - school15               1    105425 55984436 15466
## - school10               1    118914 55997926 15466
## + movie10                1     33984 55845027 15466
## - school05               1    126734 56005746 15466
## + office15               1     21304 55857707 15466
## + hospital10             1     18268 55860744 15466
## - kid10                  1    136157 56015169 15466
## + movie15                1     15636 55863376 15466
## + office05               1     14256 55864755 15466
## + bigMarket10            1      1163 55877849 15467
## + kid15                  1        75 55878936 15467
## - kid05                  1    789597 56668609 15483
## - subway05               1    886958 56765970 15486
## - subway10               1    903925 56782937 15486
## - floor                  1   1021089 56900101 15489
## - movie05                1   1368027 57247039 15498
## - hospital15             1   1388432 57267443 15499
```

```
## - subway15              1   1604790 57483801 15504
## - bigMarket05           1   1784763 57663774 15509
## - office10              1   1848634 57727646 15510
## - bigMarket15           1   2653287 58532298 15530
## - hospital05            1   2796343 58675354 15534
## - transaction_month    10   4706593 60585605 15563
## - dong                 23  15126442 71005453 15768
## - year_of_completion_f  3  21339730 77218742 15930
##
## Step:  AIC=15464.76
## unit_price ~ dong + floor + bigMarket05 + bigMarket15 + school05 +
##     school10 + school15 + subway05 + subway10 + subway15 + hospital05 +
##     hospital15 + movie05 + kid05 + kid10 + office10 + transaction_month +
##     year_of_completion_f
##
##                        Df Sum of Sq      RSS   AIC
## <none>                              55950732 15465
## + exclusive_use_area    1     71720 55879012 15465
## - school15              1    104127 56054859 15466
## - school10              1    104275 56055006 15466
## - kid10                 1    112046 56062777 15466
## + movie10               1     37063 55913669 15466
## + office15              1     32013 55918719 15466
## + hospital10            1     20500 55930232 15466
## + movie15               1     16846 55933886 15466
## + office05              1     12537 55938194 15466
## - school05              1    143882 56094613 15466
## + bigMarket10           1       559 55950173 15467
## + kid15                 1         0 55950731 15467
## - kid05                 1    763488 56714220 15482
## - subway05              1    829337 56780069 15484
## - subway10              1    877842 56828574 15485
## - floor                 1    976326 56927058 15488
## - hospital15            1   1318606 57269338 15497
## - movie05               1   1349129 57299860 15497
## - subway15              1   1586661 57537392 15503
## - bigMarket05           1   1733210 57683942 15507
## - office10              1   1803577 57754309 15509
## - bigMarket15           1   2611317 58562049 15529
## - hospital05            1   2728002 58678734 15532
## - transaction_month    10   4662039 60612771 15561
## - dong                 23  15361678 71312409 15772
## - year_of_completion_f  3  21491961 77442692 15932
```

```
##
## Call:
## lm(formula = unit_price ~ dong + floor + bigMarket05 + bigMarket15 +
##     school05 + school10 + school15 + subway05 + subway10 + subway15 +
##     hospital05 + hospital15 + movie05 + kid05 + kid10 + office10 +
##     transaction_month + year_of_completion_f, data = data_train1)
##
## Coefficients:
##             (Intercept)            dong동빙고동               dong동자동
##                 555.750                 214.387                 309.688
##              dong문배동              dong보광동              dong산천동
##                 -15.495                 378.887                -386.389
##             dong서빙고동              dong신계동              dong신창동
##                 547.168                -136.858                -315.540
##              dong용문동            dong용산동2가           dong용산동5가
##                -769.872                  25.920                -315.900
##            dong원효로1가           dong원효로2가           dong원효로4가
##                -432.411                -231.680                -122.154
##              dong이촌동              dong이태원동              dong청암동
##                 -60.282                 339.729                -300.326
##            dong한강로1가           dong한강로2가           dong한강로3가
##                  81.274                -150.889                -312.911
##              dong한남동              dong효창동              dong후암동
##                 173.574                -210.744                 398.499
##                   floor             bigMarket05             bigMarket15
##                   4.147                -134.921                  96.017
##                school05                school10                school15
##                 -19.928                 -11.901                 -10.732
##                subway05                subway10                subway15
##                 -60.501                 -39.686                  63.241
##              hospital05              hospital15                 movie05
##                  12.290                  -4.253                  12.111
##                   kid05                   kid10                office10
##                  19.154                   3.541                 -12.087
##     transaction_month02     transaction_month03     transaction_month04
##                  27.844                  50.556                  30.289
##     transaction_month05     transaction_month06     transaction_month07
##                  72.618                  89.725                 112.907
##     transaction_month08     transaction_month09     transaction_month10
##                 152.206                 173.716                 152.572
##     transaction_month11 year_of_completion_f2nd year_of_completion_f3rd
##                 240.822                  22.906                 131.764
## year_of_completion_f4th
##                 569.132
```

Linear regression prediction & RMSE calculation

```
linear_best<-lm(formula = unit_price ~ dong + floor + bigMarket05 + bigMarket15 +
    school05 + school10 + school15 + subway05 + subway10 + subway15 +
    hospital05 + hospital15 + movie05 + kid05 + kid10 + office10 +
    transaction_month + year_of_completion_f, data = data_train1)

# test data 에 적용
predict_2 <- predict(linear_best, data_test1[,-25])
summary(predict_2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   132.1   796.1   934.6   988.5  1137.5  1884.8
```

```
data_test1 %>% select(dong) %>% unique()
```

```
##           dong
## 1697    후암동
## 1699    신창동
## 1700    산천동
## 1703 원효로4가
## 1706    효창동
## 1711    문배동
## 1719    신계동
## 1726    이촌동
## 1748    한남동
## 1762  서빙고동
## 1763    보광동
## 1766 용산동2가
## 1768    동자동
## 1779    청암동
## 1792    도원동
## 1799    용문동
## 1806 한강로2가
## 1809 한강로3가
## 1876  이태원동
## 1911 원효로1가
## 2165 한강로1가
## 2504 용산동5가
## 3416  동빙고동
```

```
data_train1 %>% select(dong) %>% unique()
```

```
##           dong
## 1696    후암동
## 1698 원효로1가
## 1701    산천동
## 1702    청암동
## 1704    효창동
## 1709    도원동
## 1710    용문동
## 1712    문배동
## 1717    신계동
## 1722 한강로1가
## 1724 한강로2가
## 1725 한강로3가
## 1728    이촌동
## 1747  이태원동
## 1749    한남동
## 1761  동빙고동
## 1767    동자동
## 1780 원효로4가
## 1808 용산동5가
## 1892  서빙고동
## 1896    보광동
## 1908 용산동2가
## 1932 원효로2가
## 1934    신창동
```

```
# actual, predicted cbind

databind2 <- cbind(data_test1[,25],predict_2)
#databind2 <- cbind(data_test1[,28],predict_2)
databind2 <- as.data.frame(databind2)
summary(databind2)
```

```
##        V1           predict_2
##  Min.   : 362.5   Min.   : 132.1
##  1st Qu.: 779.3   1st Qu.: 796.1
##  Median : 937.4   Median : 934.6
##  Mean   :1016.1   Mean   : 988.5
##  3rd Qu.:1154.1   3rd Qu.:1137.5
##  Max.   :2983.9   Max.   :1884.8
```

```
# RMSE 계산
install.packages("Metrics", repos ="http://cran.us.r-project.org")
```

```
## Warning: 패키지 'Metrics'가 사용중이므로 설치되지 않을 것입니다
```

```
library(Metrics)
rmse(databind2$V1, databind2$predict_2)
```

```
## [1] 203.5205
```

Random Forest

```
install.packages("randomForest", repos ="http://cran.us.r-project.org")
```

```
## 패키지 'randomForest'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:₩Users₩LUIS₩AppData₩Local₩Temp₩Rtmpkp5M2w₩downloaded_packages
```

```
library(randomForest)
```

```
## randomForest 4.7-1.1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## 다음의 패키지를 부착합니다: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
rf.tree1 <- randomForest(unit_price~.-year, data = data_train1,
                         importance = TRUE,
                         ntree = 1000,mtry = 2)

# tree 결과
print(rf.tree1)
```

```
##
## Call:
##  randomForest(formula = unit_price ~ . - year, data = data_train1,      importance = TRUE, n
tree = 1000, mtry = 2)
##                Type of random forest: regression
##                      Number of trees: 1000
## No. of variables tried at each split: 2
##
##           Mean of squared residuals: 15296.22
##                     % Var explained: 86.99
```

```
summary(rf.tree1)
```

```
##                Length Class  Mode
## call              6   -none- call
## type              1   -none- character
## predicted      1455   -none- numeric
## mse            1000   -none- numeric
## rsq            1000   -none- numeric
## oob.times      1455   -none- numeric
## importance       52   -none- numeric
## importanceSD     26   -none- numeric
## localImportance   0   -none- NULL
## proximity         0   -none- NULL
## ntree             1   -none- numeric
## mtry              1   -none- numeric
## forest           11   -none- list
## coefs             0   -none- NULL
## y              1455   -none- numeric
## test              0   -none- NULL
## inbag             0   -none- NULL
## terms             3   terms  call
```

```
install.packages("rpart.plot", repos ="http://cran.us.r-project.org")
```

```
## Warning: 패키지 'rpart.plot'가 사용중이므로 설치되지 않을 것입니다
```

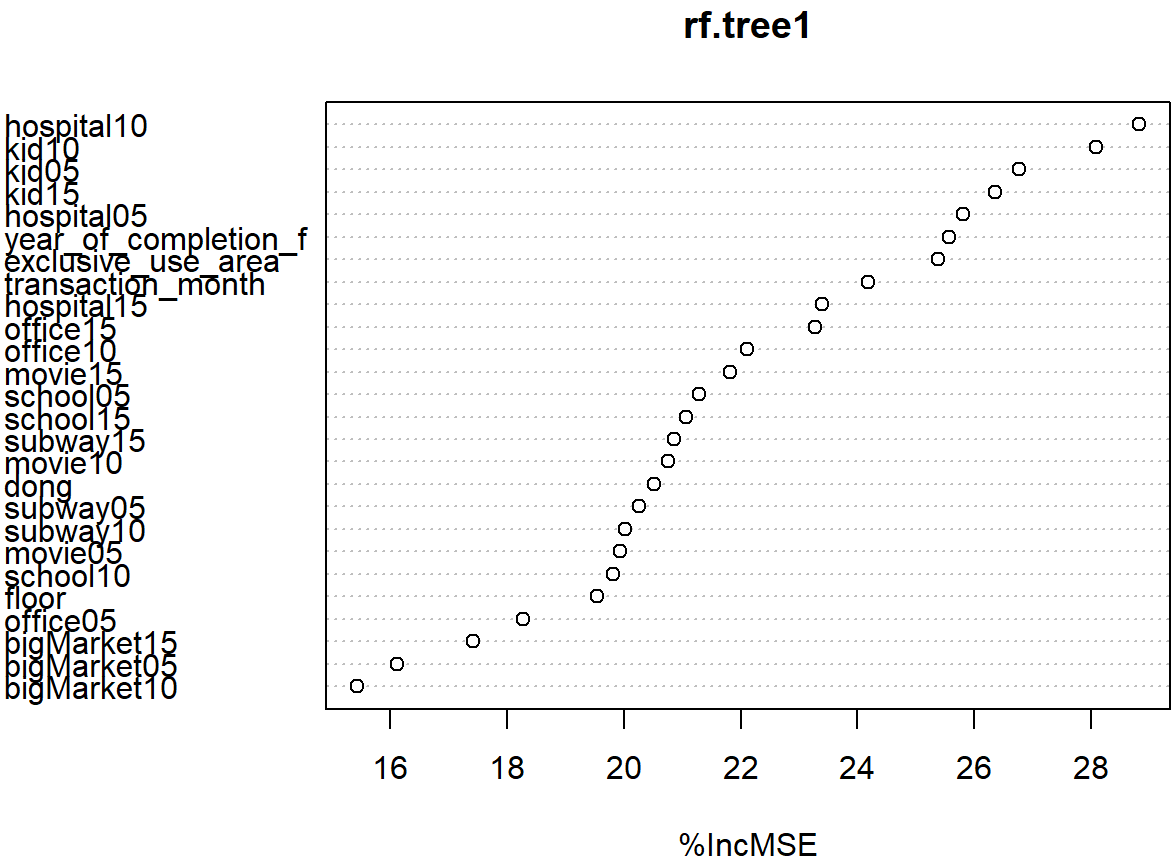
```
library(rpart.plot)

importance(rf.tree1)
```

```
##                       %IncMSE IncNodePurity
## dong                 20.51053       6990538
## exclusive_use_area   25.37573      12063813
## floor                19.54729       2557191
## bigMarket05          16.11794       2865625
## bigMarket10          15.43517       1482235
## bigMarket15          17.41190       3612799
## school05             21.29455       2724969
## school10             19.81492       3033173
## school15             21.06584       6193147
## subway05             20.26130       2863488
## subway10             20.01496       2247810
## subway15             20.85838       5005737
## hospital05           25.80644       8773606
## hospital10           28.82043       5653081
## hospital15           23.39589       7055012
## movie05              19.93241       9177406
## movie10              20.75895       9203158
## movie15              21.82524       7437700
## kid05                26.77038       4268844
## kid10                28.08596       6598606
## kid15                26.36561       8806458
## office05             18.26785       5449023
## office10             22.10263       8663977
## office15             23.28420       7712912
## transaction_month    24.17651       2897596
## year_of_completion_f 25.57604       8633188
```

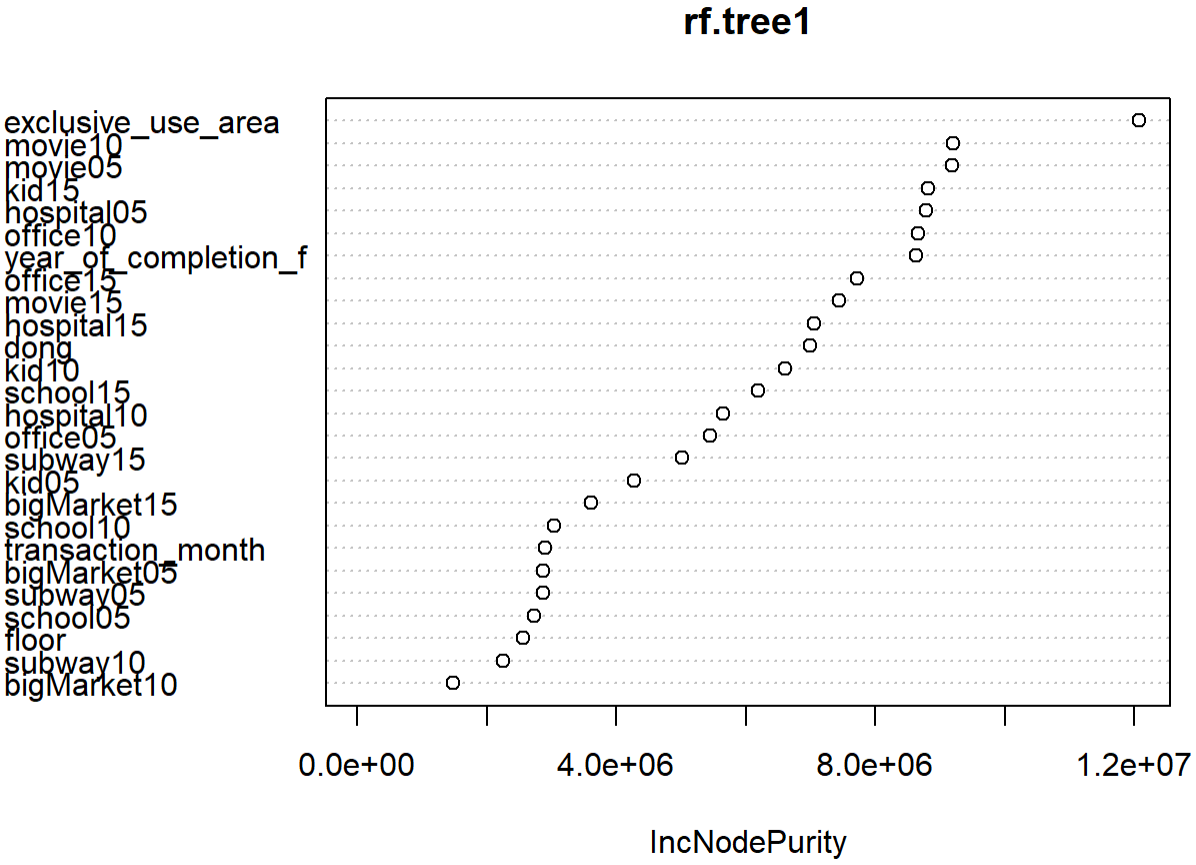
```
importance(rf.tree1, type = 1)
```

```
##                       %IncMSE
## dong                 20.51053
## exclusive_use_area   25.37573
## floor                19.54729
## bigMarket05          16.11794
## bigMarket10          15.43517
## bigMarket15          17.41190
## school05             21.29455
## school10             19.81492
## school15             21.06584
## subway05             20.26130
## subway10             20.01496
## subway15             20.85838
## hospital05           25.80644
## hospital10           28.82043
## hospital15           23.39589
## movie05              19.93241
## movie10              20.75895
## movie15              21.82524
## kid05                26.77038
## kid10                28.08596
## kid15                26.36561
## office05             18.26785
## office10             22.10263
## office15             23.28420
## transaction_month    24.17651
## year_of_completion_f 25.57604
```

```
varImpPlot(rf.tree1, type = 1)
```




```
varImpPlot(rf.tree1, type = 2)
```

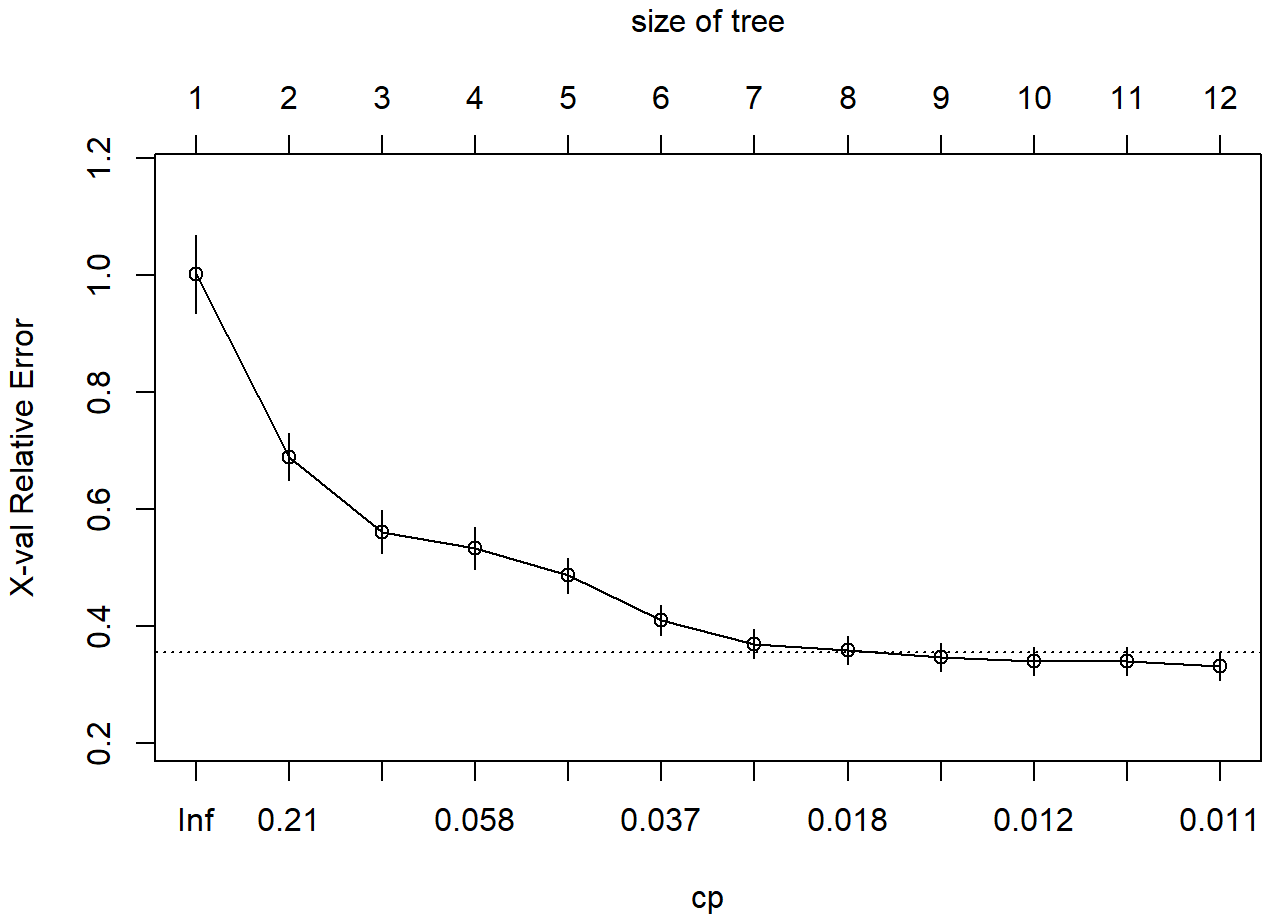


Random Forest parameter tuning

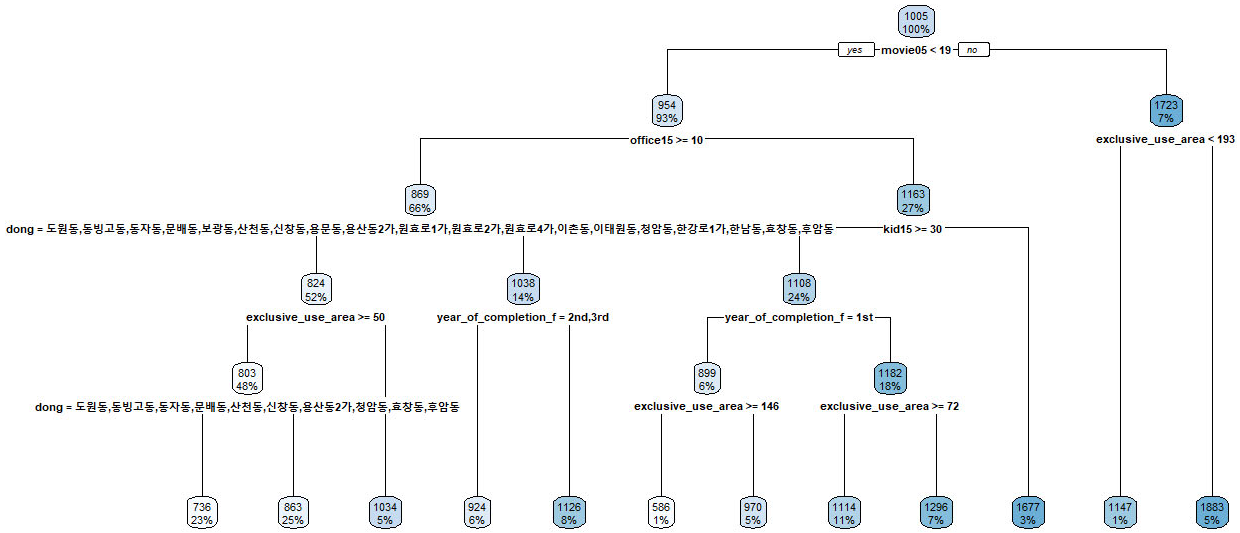
```
printcp(tree1)
```

```
##
## Regression tree:
## rpart(formula = unit_price ~ . - year, data = data_train1, method = "anova",
##     control = rpart.control(minsplit = 50, maxdepth = 5))
##
## Variables actually used in tree construction:
## [1] dong                 exclusive_use_area   kid15
## [4] movie05              office15             year_of_completion_f
##
## Root node error: 171065491/1455 = 117571
##
## n= 1455
##
##          CP nsplit rel error  xerror     xstd
## 1  0.313548      0   1.00000 1.00247 0.066323
## 2  0.141099      1   0.68645 0.68971 0.040205
## 3  0.065065      2   0.54535 0.56204 0.036761
## 4  0.052111      3   0.48029 0.53321 0.035824
## 5  0.043217      4   0.42818 0.48715 0.029294
## 6  0.032058      5   0.38496 0.41147 0.025583
## 7  0.019562      6   0.35290 0.37061 0.024297
## 8  0.016441      7   0.33334 0.35935 0.024149
## 9  0.011927      8   0.31690 0.34709 0.023986
## 10 0.011920      9   0.30497 0.34106 0.023935
## 11 0.011826     10   0.29305 0.34106 0.023935
## 12 0.010000     11   0.28123 0.33286 0.023818
```

```
plotcp(tree1)
```




```
tree1 <- prune(tree1, cp= tree1$cptable[which.min(tree1$cptable[, "xerror"]), "CP"])

rpart.plot(tree1)
```



Random Forest prediction & RMSE calculation

```
# test data 에 적용

predict_3 <- predict(rf.tree1, data_test1)
summary(predict_3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   499.1   790.3   939.1  1007.6  1134.2  2038.3
```

```
# actual, predicted cbind

databind3 <- cbind(data_test1[,25],predict_3)
databind3 <- as.data.frame(databind3)
summary(databind3)
```

```
##        V1           predict_3
##  Min.   : 362.5   Min.   : 499.1
##  1st Qu.: 779.3   1st Qu.: 790.3
##  Median : 937.4   Median : 939.1
##  Mean   :1016.1   Mean   :1007.6
##  3rd Qu.:1154.1   3rd Qu.:1134.2
##  Max.   :2983.9   Max.   :2038.3
```

```
# RMSE 계산
install.packages("Metrics", repos ="http://cran.us.r-project.org")
```

```
## Warning: 패키지 'Metrics'가 사용중이므로 설치되지 않을 것입니다
```

```
library(Metrics)
rmse(databind3$V1, databind3$predict_3)
```

```
## [1] 129.1248
```

Gradient Boost Model

```
install.packages("gbm", repos ="http://cran.us.r-project.org")
```

```
## 패키지 'gbm'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:₩Users₩LUIS₩AppData₩Local₩Temp₩Rtmpkp5M2w₩downloaded_packages
```

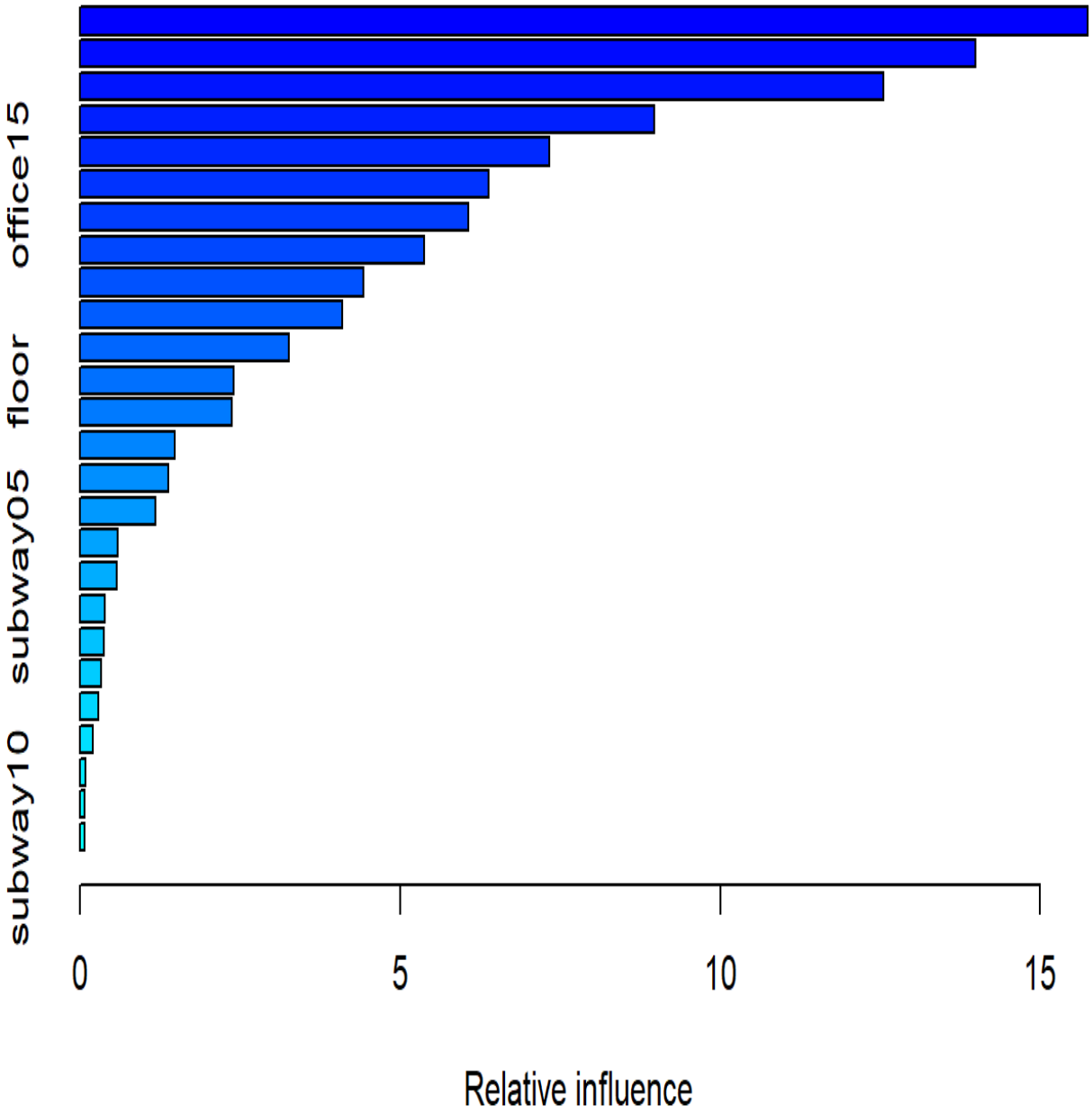
```
library(gbm)
```

```
## Loaded gbm 2.1.8.1
```

```
gbm.tree1 <- gbm(unit_price~.-year, data = data_train1, distribution = "gaussian",
                 n.trees = 1000, shrinkage = 0.01, interaction.depth = 4)

# tree 결과
print(gbm.tree1)
```

```
## gbm(formula = unit_price ~ . - year, distribution = "gaussian",
##     data = data_train1, n.trees = 1000, interaction.depth = 4,
##     shrinkage = 0.01)
## A gradient boosted model with gaussian loss function.
## 1000 iterations were performed.
## There were 26 predictors of which 26 had non-zero influence.
```

```
summary(gbm.tree1)
```


office15
floor
subway05
subway10
0
5
10
15
Relative influence

```
##                                     var     rel.inf
## exclusive_use_area     exclusive_use_area 15.73962745
## movie05                           movie05 13.98291750
## dong                                 dong 12.55133941
## year_of_completion_f year_of_completion_f  8.97750323
## hospital05                     hospital05  7.32999233
## office15                         office15  6.38301734
## movie10                           movie10  6.06380874
## kid15                               kid15  5.38270871
## transaction_month       transaction_month  4.42863449
## school15                         school15  4.09665109
## movie15                           movie15  3.25968555
## floor                               floor  2.39903853
## hospital10                     hospital10  2.37468390
## office10                         office10  1.48519737
## kid10                               kid10  1.38271702
## hospital15                     hospital15  1.17466575
## kid05                               kid05  0.58419255
## subway05                         subway05  0.57207630
## school05                         school05  0.38597057
## school10                         school10  0.38031156
## office05                         office05  0.33579570
## subway15                         subway15  0.28468897
## bigMarket05                   bigMarket05  0.20766551
## bigMarket10                   bigMarket10  0.08411173
## bigMarket15                   bigMarket15  0.07942797
## subway10                         subway10  0.07357073
```

Gradient Boost Model parameter tuning

```
# printcp(tree1)
# plotcp(tree1)
# tree1 <- prune(tree1, cp= tree1$cptable[which.min(tree1$cptable[, "xerror"]), "CP"])
#
# rpart.plot(tree1)
```

##Gradient Boost Model prediction & RMSE calculation

```
# test data 에 적용
predict_4 <- predict.gbm(object = gbm.tree1,
                         newdata = data_test1,
                         n.trees = 1000,
                         type = "response")
summary(predict_4)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   484.4   800.9   913.3  1001.5  1137.1  2304.0
```

```
# actual, predicted cbind

databind4 <- cbind(data_test1[,25],predict_4)
databind4 <- as.data.frame(databind4)
summary(databind4)
```

```
##        V1           predict_4
##  Min.   : 362.5   Min.   : 484.4
##  1st Qu.: 779.3   1st Qu.: 800.9
##  Median : 937.4   Median : 913.3
##  Mean   :1016.1   Mean   :1001.5
##  3rd Qu.:1154.1   3rd Qu.:1137.1
##  Max.   :2983.9   Max.   :2304.0
```

```
# RMSE 계산
install.packages("Metrics", repos ="http://cran.us.r-project.org")
```

```
## Warning: 패키지 'Metrics'가 사용중이므로 설치되지 않을 것입니다
```

```
library(Metrics)
rmse(databind4$V1, databind4$predict_4)
```

```
## [1] 124.4346
```