

2015-05-18.R

MaciejBerezewicz

Tue May 19 10:07:19 2015

```
library(spgwr)
```

```
## Loading required package: sp
## NOTE: This package does not constitute approval of GWR
## as a method of spatial analysis; see example(gwr)
```

```
library(foreign)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##   filter
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggmap)
```

```
## Loading required package: ggplot2
```

```
library(ggplot2)
library(AICcmodavg)
```

```
baltimore <- read.dbf('~/Documents/Projects/RProjects/Dydaktyka/REMA/2014-2015/datasets/baltim.dbf')
```

```
head(baltimore)
```

```
##   STATION PRICE NROOM DWELL NBATH PATIO FIREPL AC BMENT NSTOR GAR AGE
## 1      1  47.0    4     0   1.0    0     0  0     2     3  0 148
## 2      2 113.0    7     1   2.5    1     1  1     2     2  2   9
## 3      3 165.0    7     1   2.5    1     1  0     3     2  2  23
## 4      4 104.3    7     1   2.5    1     1  1     2     2  2   5
## 5      5  62.5    7     1   1.5    1     1  0     2     2  0  19
## 6      6  70.0    6     1   2.5    1     1  0     3     3  1  20
##   CITCOU LOTSZ  SQFT   X   Y
## 1      0   5.70 11.25 907 534
## 2      1 279.51 28.92 922 574
## 3      1  70.64 30.62 920 581
## 4      1 174.63 26.12 923 578
## 5      1 107.80 22.04 918 574
## 6      1 139.64 39.42 900 577
```

```
simple_model <- lm(PRICE ~ NROOM + NBATH + GAR + AGE + SQFT, data = baltimore)
summary(simple_model)
```

```
##
## Call:
## lm(formula = PRICE ~ NROOM + NBATH + GAR + AGE + SQFT, data = baltimore)
##
## Residuals:
```

| | Min | 1Q | Median | 3Q | Max |
|--|---------|--------|--------|-------|--------|
| | -66.434 | -9.353 | -2.320 | 7.112 | 78.301 |

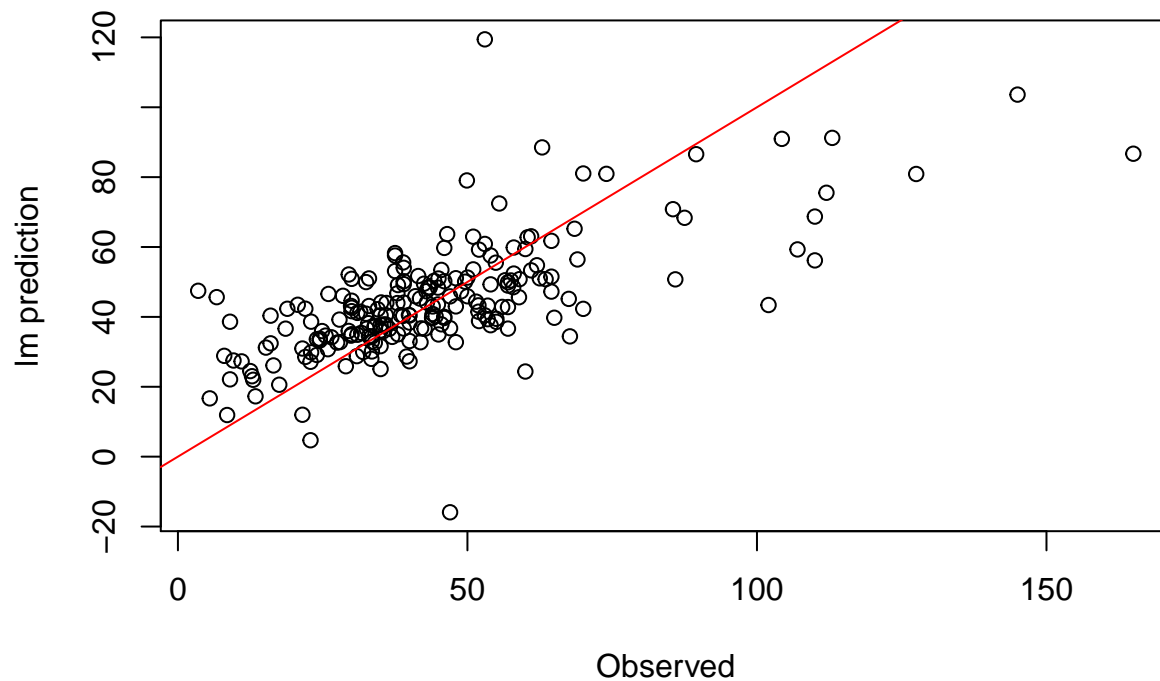
```
##
## Coefficients:
```

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 24.46638 | 6.16030 | 3.972 | 9.88e-05 *** |
| NROOM | 1.07133 | 1.47598 | 0.726 | 0.46876 |
| NBATH | 7.85676 | 2.41519 | 3.253 | 0.00134 ** |
| GAR | 11.79618 | 2.19983 | 5.362 | 2.21e-07 *** |
| AGE | -0.40667 | 0.06241 | -6.516 | 5.49e-10 *** |
| SQFT | 0.68101 | 0.21755 | 3.130 | 0.00200 ** |

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.38 on 205 degrees of freedom
## Multiple R-squared:  0.4707, Adjusted R-squared:  0.4578
## F-statistic: 36.46 on 5 and 205 DF,  p-value: < 2.2e-16
```

```
plot(x = baltimore$PRICE,
     y = simple_model$fitted.values,
     main = 'Comparison of observed values (X) and prediction from LW (Y)',
     xlab = 'Observed',
     ylab = 'lm prediction')
abline(a=0,b=1,col='red')
```

Comparison of observed values (X) and prediction from LW (Y)



```
### setting the bandwidth (b)
```

```
bandwidth <- gwr.sel(  
  formula = PRICE ~ NROOM + NBATH +  
    GAR + AGE + SQFT,  
  data = baltimore,  
  coords = cbind(  
    baltimore$X,  
    baltimore$Y))
```

```
## Bandwidth: 56.69026 CV score: 68902.3  
## Bandwidth: 91.63518 CV score: 70064.57  
## Bandwidth: 35.0931 CV score: 66010.8  
## Bandwidth: 21.74533 CV score: 60793.44  
## Bandwidth: 13.49595 CV score: 56642.9  
## Bandwidth: 8.397555 CV score: 57913.94  
## Bandwidth: 13.15798 CV score: 56570.02  
## Bandwidth: 12.22299 CV score: 56468.09  
## Bandwidth: 10.7618 CV score: 56656.8  
## Bandwidth: 12.03985 CV score: 56467.25  
## Bandwidth: 12.10672 CV score: 56466.78  
## Bandwidth: 12.10827 CV score: 56466.78  
## Bandwidth: 12.10806 CV score: 56466.78  
## Bandwidth: 12.1081 CV score: 56466.78  
## Bandwidth: 12.10802 CV score: 56466.78  
## Bandwidth: 12.10806 CV score: 56466.78
```

```
### estimate GWR model
```

```
model <- gwr(  
  formula = PRICE ~ NROOM + NBATH +  
    GAR + AGE + SQFT,  
  data = baltimore,  
  coords = cbind(  
    baltimore$X,  
    baltimore$Y),  
  bandwidth = bandwidth)  
  
print(model)
```

```
## Call:
```

```
## gwr(formula = PRICE ~ NROOM + NBATH + GAR + AGE + SQFT, data = baltimore,  
##      coords = cbind(baltimore$X, baltimore$Y), bandwidth = bandwidth)  
## Kernel function: gwr.Gauss  
## Fixed bandwidth: 12.10806  
## Summary of GWR coefficient estimates at data points:  
##           Min.   1st Qu.   Median   3rd Qu.   Max.   Global  
## X.Intercept. -0.24240  14.89000  22.31000  32.92000  56.56000 24.4664  
## NROOM        -2.67800   0.68340   1.76900   2.91200   5.84300  1.0713  
## NBATH       -12.06000   3.11300   5.02700   8.33600  20.82000  7.8568  
## GAR         -8.71100   2.30700   5.00800   9.12500  21.62000 11.7962  
## AGE         -0.88270  -0.45840  -0.34600  -0.18540   0.01957 -0.4067  
## SQFT        -0.31790   0.32480   0.52580   0.68890   1.59600  0.6810
```

```
model <- gwr(  
  formula = PRICE ~ NROOM + NBATH +  
    GAR + AGE + SQFT,  
  data = baltimore,  
  coords = cbind(  
    baltimore$X,  
    baltimore$Y),  
  bandwidth = bandwidth,  
  hatmatrix = TRUE, ## hat matrix  
  se.fit = TRUE, ## standard errors  
  cl = 8, ## parallel computing  
  predictions = TRUE) ## save predictions  
  
print(model)
```

```
## Call:
```

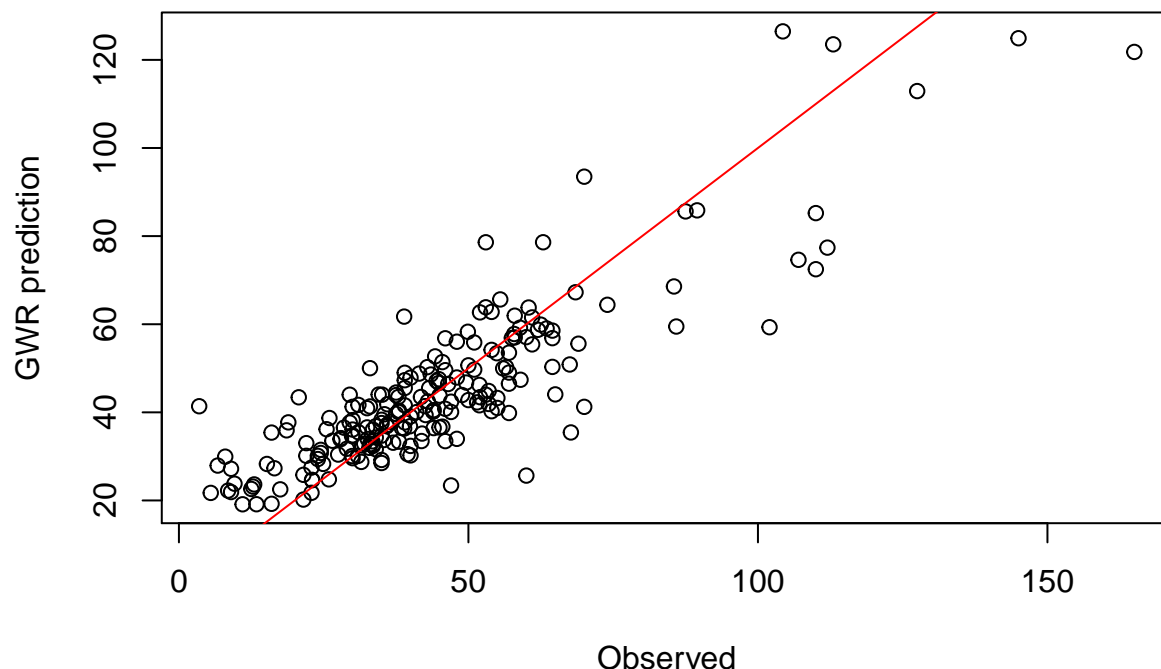
```
## gwr(formula = PRICE ~ NROOM + NBATH + GAR + AGE + SQFT, data = baltimore,  
##      coords = cbind(baltimore$X, baltimore$Y), bandwidth = bandwidth,  
##      hatmatrix = TRUE, se.fit = TRUE, cl = 8, predictions = TRUE)  
## Kernel function: gwr.Gauss  
## Fixed bandwidth: 12.10806  
## Summary of GWR coefficient estimates at data points:  
##           Min.   1st Qu.   Median   3rd Qu.   Max.   Global  
## X.Intercept. -0.24240  14.89000  22.31000  32.92000  56.56000 24.4664  
## NROOM        -2.67800   0.68340   1.76900   2.91200   5.84300  1.0713  
## NBATH       -12.06000   3.11300   5.02700   8.33600  20.82000  7.8568
```

```
## GAR          -8.71100   2.30700   5.00800   9.12500  21.62000  11.7962
## AGE          -0.88270  -0.45840  -0.34600  -0.18540   0.01957  -0.4067
## SQFT         -0.31790   0.32480   0.52580   0.68890   1.59600   0.6810
## Number of data points: 211
## Effective number of parameters (residual: 2traceS - traceS'S): 59.65483
## Effective degrees of freedom (residual: 2traceS - traceS'S): 151.3452
## Sigma (residual: 2traceS - traceS'S): 14.01235
## Effective number of parameters (model: traceS): 44.09774
## Effective degrees of freedom (model: traceS): 166.9023
## Sigma (model: traceS): 13.34333
## Sigma (ML): 11.86736
## AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 1758.141
## AIC (GWR p. 96, eq. 4.22): 1686.83
## Residual sum of squares: 29716.03
## Quasi-global R2: 0.7460643
```

```
## compare results
```

```
plot(x = baltimore$PRICE,
     y = model$SDF@data$pred,
     main = 'Comparison of observed values (X) and prediction from GWR (Y)',
     xlab = 'Observed',
     ylab = 'GWR prediction')
abline(a=0,b=1,col='red')
```

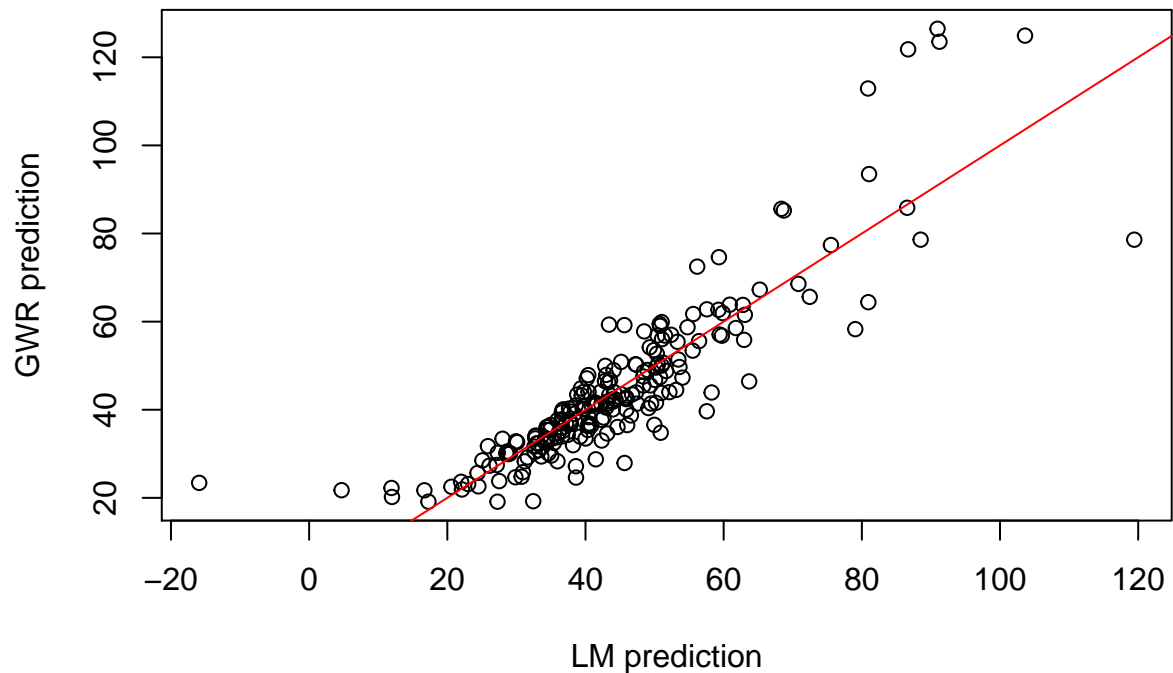
Comparison of observed values (X) and prediction from GWR (Y)



```
plot(x = simple_model$fitted.values,
     y = model$SDF@data$pred,
     main = 'Comparison of prediction from LM (X) and GWR (Y)',
     xlab = 'LM prediction',
```

```
ylab = 'GWR prediction')
abline(a=0,b=1,col='red')
```

Comparison of prediction from LM (X) and GWR (Y)



```
### visualisation
```

```
str(model,1)
```

```
## List of 12
## $ SDF      :Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
## $ lhat     : num [1:211, 1:211] 4.80e-01 -8.17e-04 5.78e-05 -3.50e-04 -5.09e-04 ...
## $ lm       :List of 11
## $ results  :List of 14
## $ bandwidth: num 12.1
## $ adapt    : NULL
## $ hatmatrix: logi TRUE
## $ gweight  : chr "gwr.Gauss"
## $ gTSS     : num 117022
## $ this.call: language gwr(formula = PRICE ~ NROOM + NBATH + GAR + AGE + SQFT, data = baltimore,
## $ fp.given : logi FALSE
## $ timings  : num [1:6, 1:2] 0.002 0.281 0.019 0.014 0 ...
## ..- attr(*, "dimnames")=List of 2
## - attr(*, "class")= chr "gwr"
```

```
str(model$SDF,2)
```

```
## Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
```

```
## ..@ data      :'data.frame':  211 obs. of  24 variables:
## ..@ coords.nrs : num(0)
## ..@ coords    : num [1:211, 1:2] 907 922 920 923 918 900 918 907 918 897 ...
## .. ..- attr(*, "dimnames")=List of 2
## ..@ bbox      : num [1:2, 1:2] 860 506 988 581
## .. ..- attr(*, "dimnames")=List of 2
## ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
```

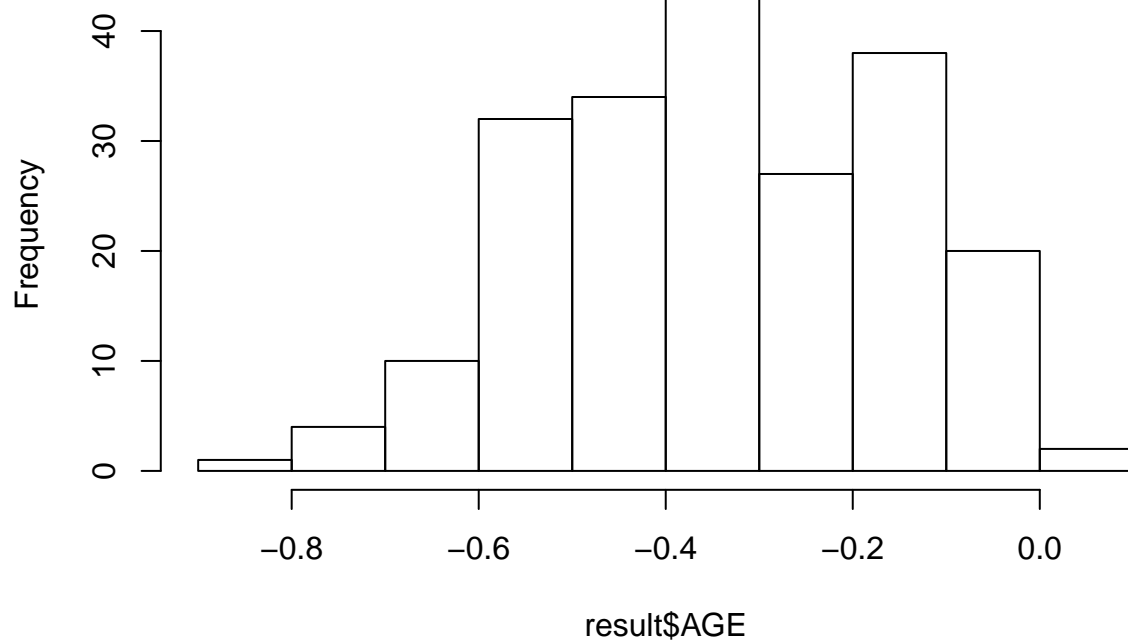
```
str(model$SDF@data,1)
```

```
## 'data.frame':  211 obs. of  24 variables:
## $ sum.w      : num  31.8 23.8 15.9 19.4 23.5 ...
## $ (Intercept) : num  4.78 11.29 2.96 6.48 13.35 ...
## $ NROOM      : num  2.03 4.56 4.52 4.74 4.03 ...
## $ NBATH      : num  3.97 12.96 18.32 14.74 14.87 ...
## $ GAR        : num  5.52 21.08 21.53 21.62 20.5 ...
## $ AGE        : num  0.0196 -0.5409 -0.3802 -0.4627 -0.5484 ...
## $ SQFT       : num  0.323 0.368 0.232 0.345 0.293 ...
## $ (Intercept)_se : num  8.26 12.62 14.62 13.94 11.8 ...
## $ NROOM_se     : num  2.07 2.69 3.09 2.94 2.59 ...
## $ NBATH_se     : num  3.56 4.37 4.85 4.62 4.28 ...
## $ GAR_se       : num  3.57 2.93 3.1 3.12 2.81 ...
## $ AGE_se       : num  0.076 0.13 0.156 0.148 0.121 ...
## $ SQFT_se      : num  0.343 0.343 0.363 0.369 0.315 ...
## $ gwr.e        : num  23.6 -10.5 43.2 -22.2 2.6 ...
## $ pred         : num  23.4 123.5 121.8 126.5 59.9 ...
## $ pred.se      : num  8.27 5.94 5.6 6.59 4.91 ...
## $ localR2      : num  0.442 0.832 0.822 0.832 0.823 ...
## $ (Intercept)_se_EDF: num  8.67 13.25 15.35 14.64 12.39 ...
## $ NROOM_se_EDF   : num  2.18 2.82 3.24 3.09 2.71 ...
## $ NBATH_se_EDF   : num  3.74 4.58 5.1 4.86 4.49 ...
## $ GAR_se_EDF     : num  3.74 3.08 3.25 3.27 2.95 ...
## $ AGE_se_EDF     : num  0.0798 0.1363 0.1637 0.1556 0.1266 ...
## $ SQFT_se_EDF    : num  0.36 0.36 0.381 0.387 0.331 ...
## $ pred.se_EDF    : num  8.69 6.23 5.88 6.92 5.15 ...
```

```
result <- model$SDF@data
```

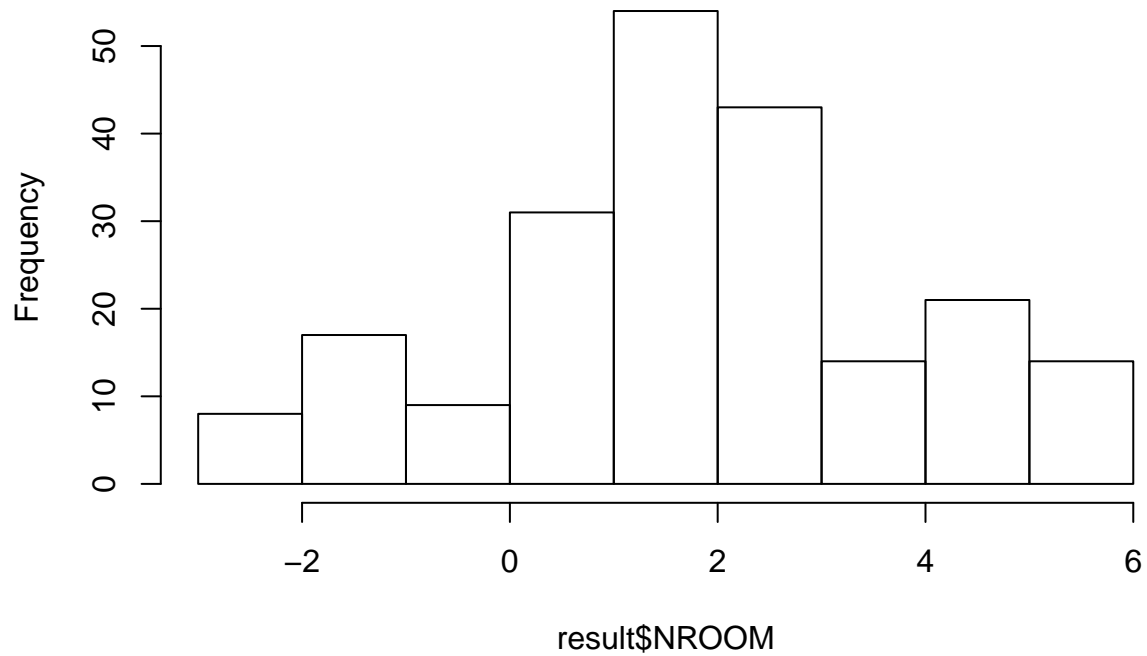
```
hist(result$AGE)
```

Histogram of result\$AGE



```
hist(result$NROOM)
```

Histogram of result\$NROOM

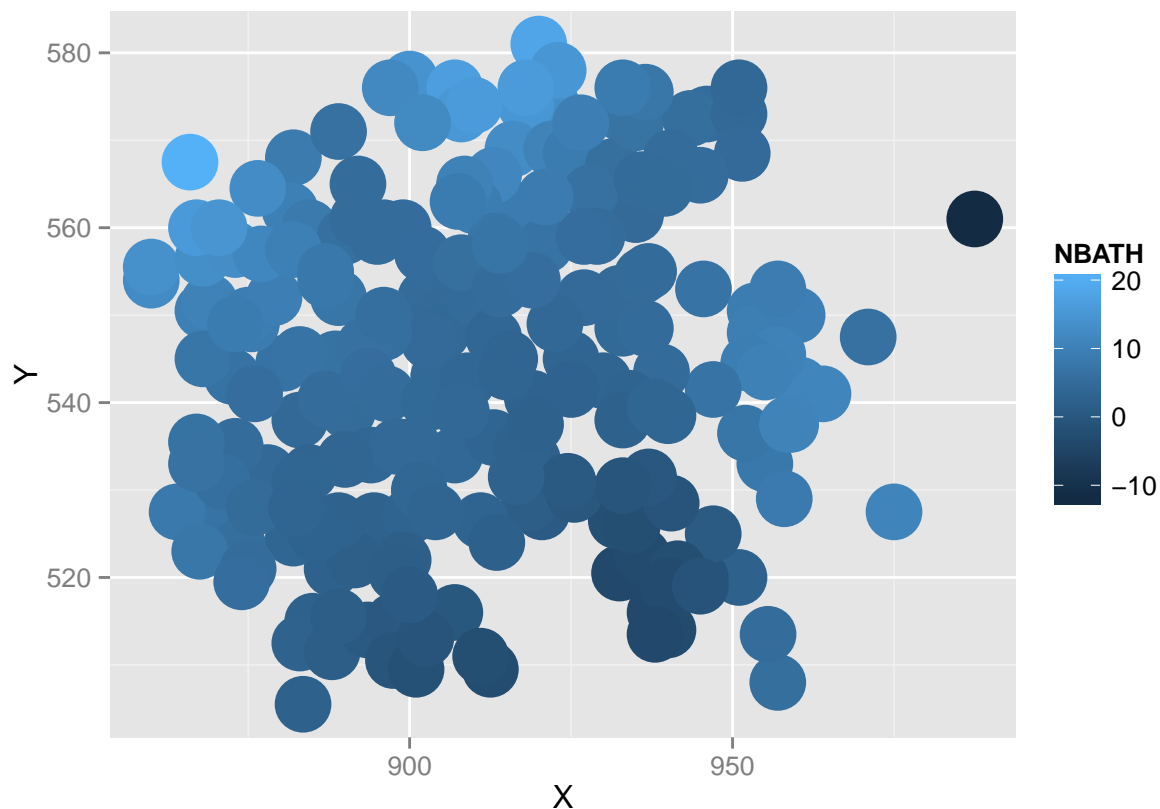



```

### plotting results
result$X <- baltimore$X
result$Y <- baltimore$Y

ggplot(
  data = result,
  aes(x = X,
      y = Y,
      colour = NBATH)) +
  geom_point(size=10)

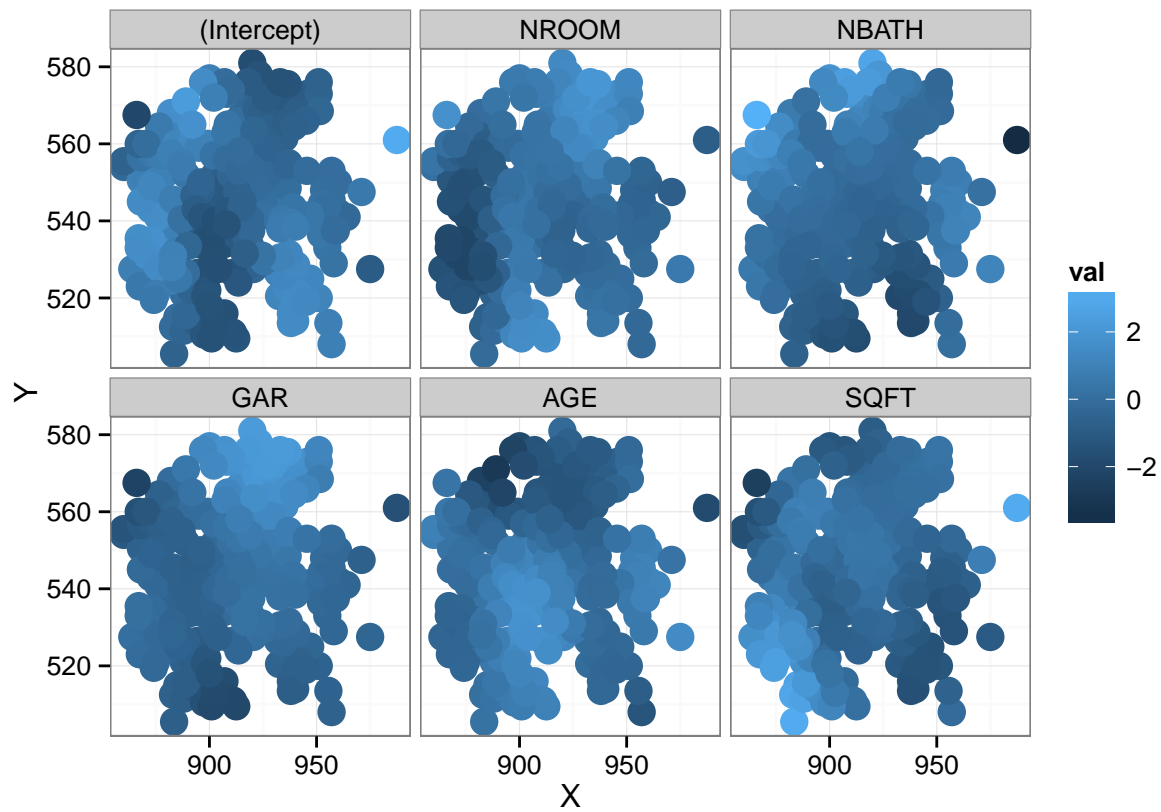
```



```

result %>%
  select(`(Intercept)`:SQFT,X,Y) %>%
  gather(var,val,-X,-Y) %>%
  group_by(var) %>%
  mutate(val = scale(val)) %>%
  ggplot(data = .,
    aes(x = X,
        y = Y,
        colour = val)) +
  geom_point(size = 5) +
  facet_wrap(~var) +
  theme_bw()

```



```
### how much did we gain?
cat('LM AICc:', AICc(simple_model),
    'GWR AICc:', model$results$AICc)
```

```
## LM AICc: 1812.265 GWR AICc: 1769.642
```