Final Project

Jillian Warburton and Mary Ebbert

2023-04-26

Section 0: Executive Summary

From this analysis, we determined that the home characteristics explain home appraisal price fairly well, as evidenced by the fit measures we explored. We created a model that accounted for the variability and interdependence of the characteristics. The following factors, on average, lead to higher home appraisal prices: larger general living area, recent remodeling, central air, larger garage capacity, and having a style of house that was not 1.5 stories with a finished basement. We also determined that the home appraisal price increased with the general living area of a home. Finally, we predicted the home appraisal prices of 52 homes.

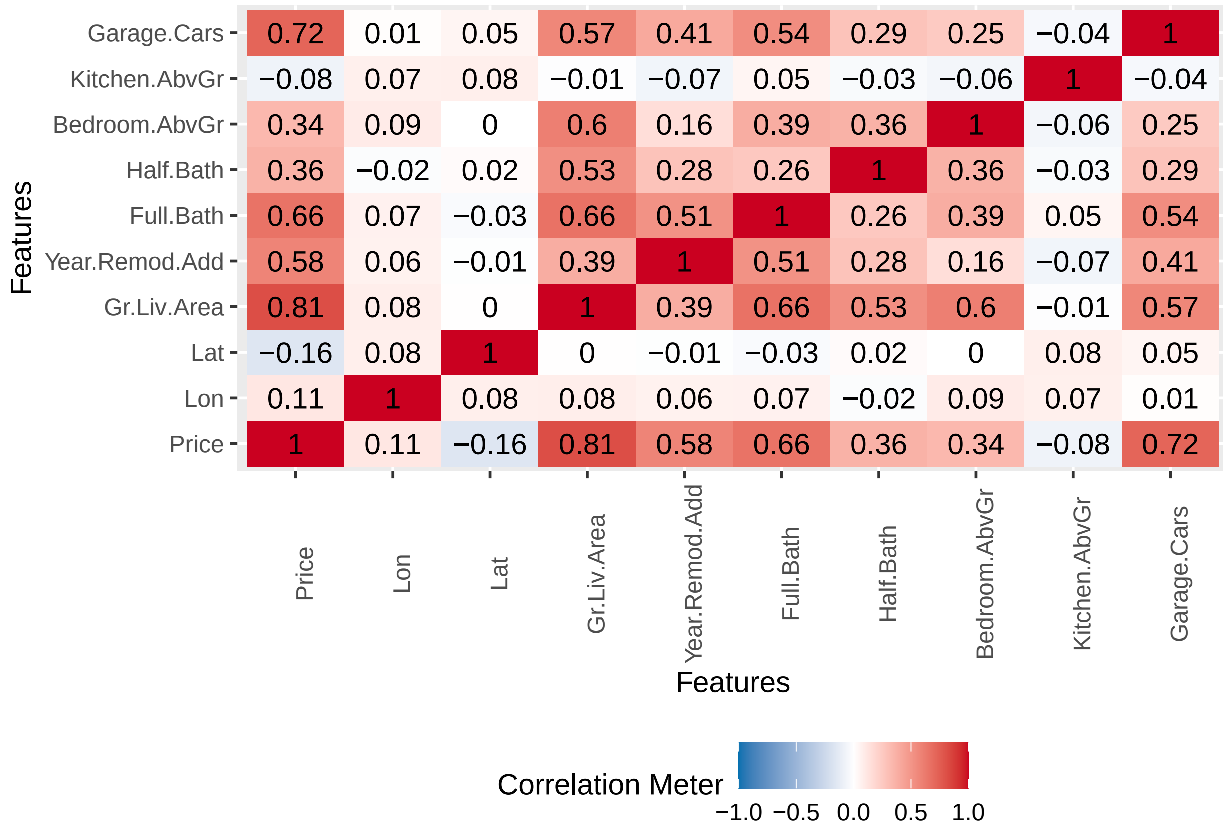
Section 1: Introduction and Problem Background

An appraisal of a house's value is necessary for mortgages as it is used as collateral against the loan. A house appraisal is an estimate of a house's market value that is meant to be unbiased, fair, and conducted by an independent third party. This analysis attempts to provide a statistical method for determining the value of a house based on the following criteria using data from Ames, IA:

Variable Name	Description
Price	Sale price of the home
Lon	A transformed measure of longitude (for privacy reasons)
Lat	A transformed measure of latitude (for privacy reasons)
Gr.Liv.Area	Above ground living area in square feet
House.style	Style of dwelling (1 Story, 2 Stories, Split Level)
Year.Remod.Add	Remodel date (if no remodel than original construction date)
Central.Air	Include central air
Full.Bath	Number of full bathrooms above ground
Half.Bath	Number of half bathrooms above ground
Bedroom.AbvGr	Number of bedrooms above ground
Kitchen.AbvGr	Number of kitchens above ground
Garage.Cars	Size of garage in car capacity

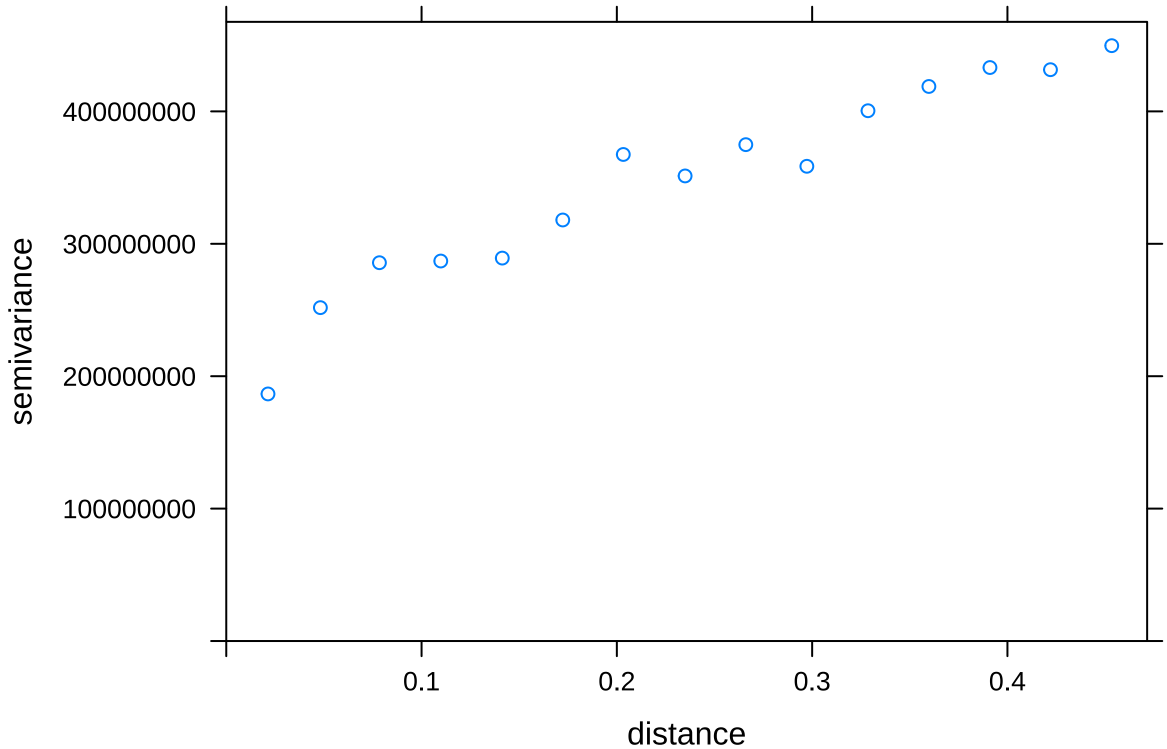
The goals with this analysis are to answer the following questions:
1. How well do the home characteristics explain sale price?
2. What factors increase the sale price of a home?
3. Does the variability of sale price increase with the size of the home (as given by living area)?
4. What is your predicted/appraised sale price for the homes in the dataset that do not have a sale price?

The following graphics summarize the data used in this analysis:

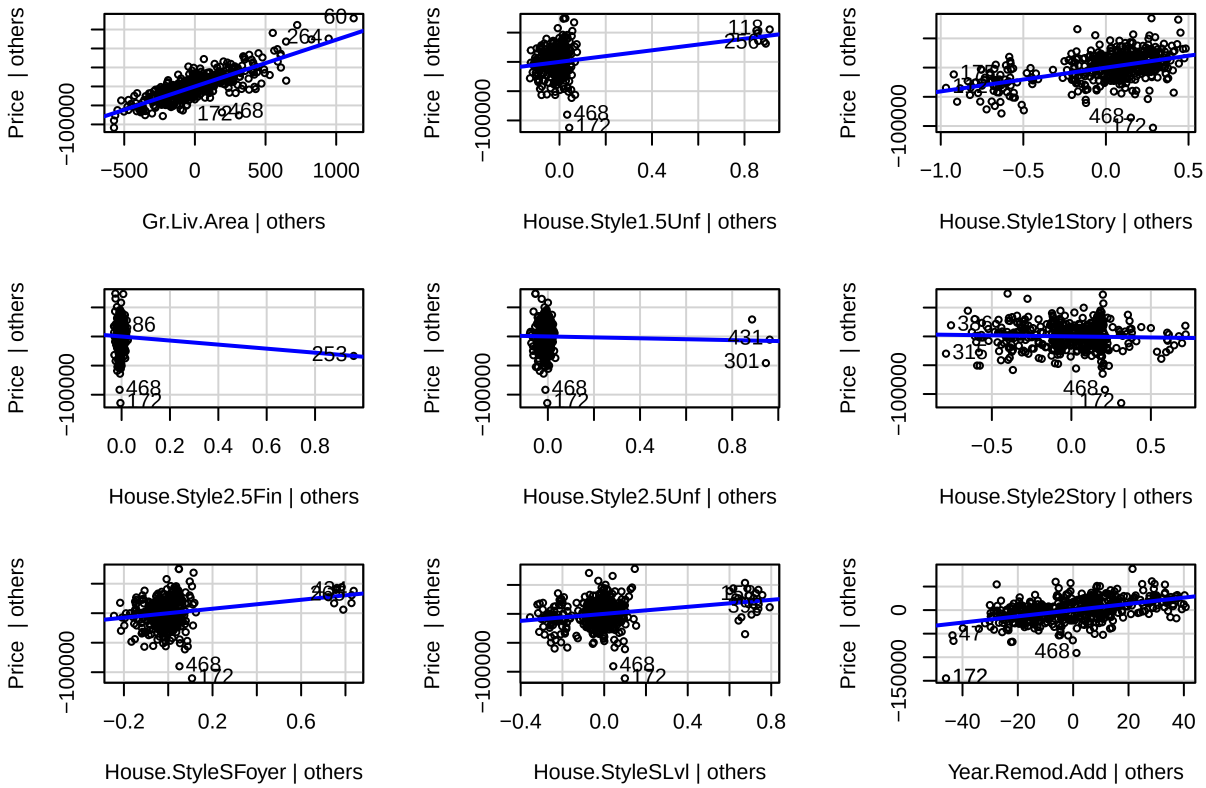


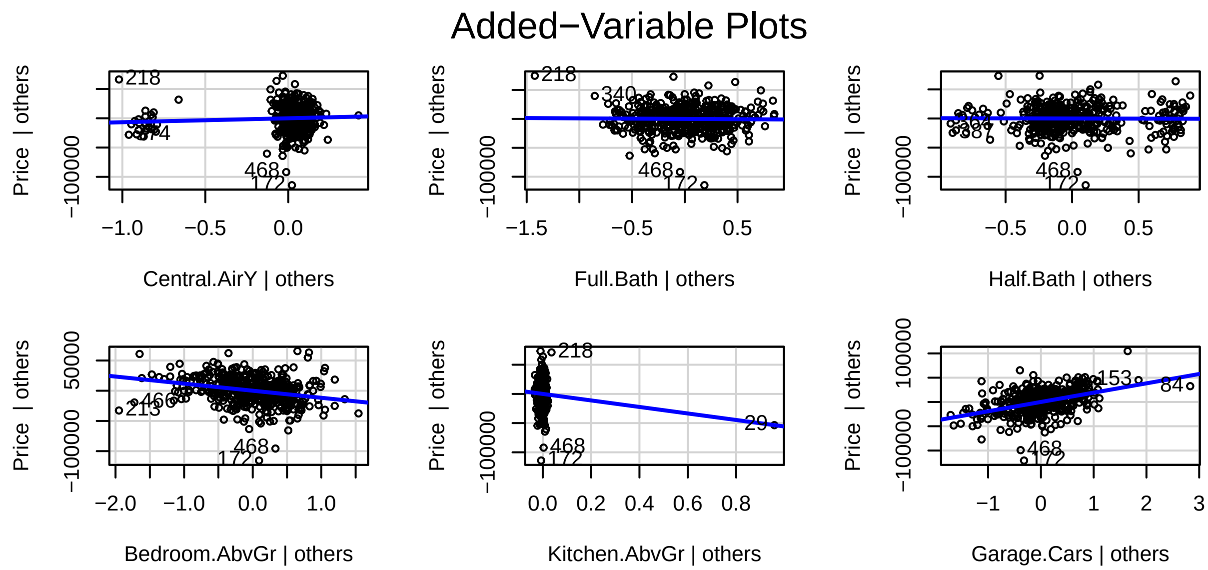
This correlation plot shows that there are several features that are strongly to moderately correlated with the house sale price. `Gr.Liv.Area` and `Garage.Cars` are the most strongly and positively correlated with the house price. `Full.Bath`, `Year.Remod.Add`, `Half.Bath` and `Bedroom.AbvGr` are also moderately positively correlated with the house price.

This plot attempts to map the house prices based on their transformed longitude and latitude. It is clear that there is some spatial correlation which will need to be accounted for. This is made increasingly more obvious by the increasing (and not flat) variogram:

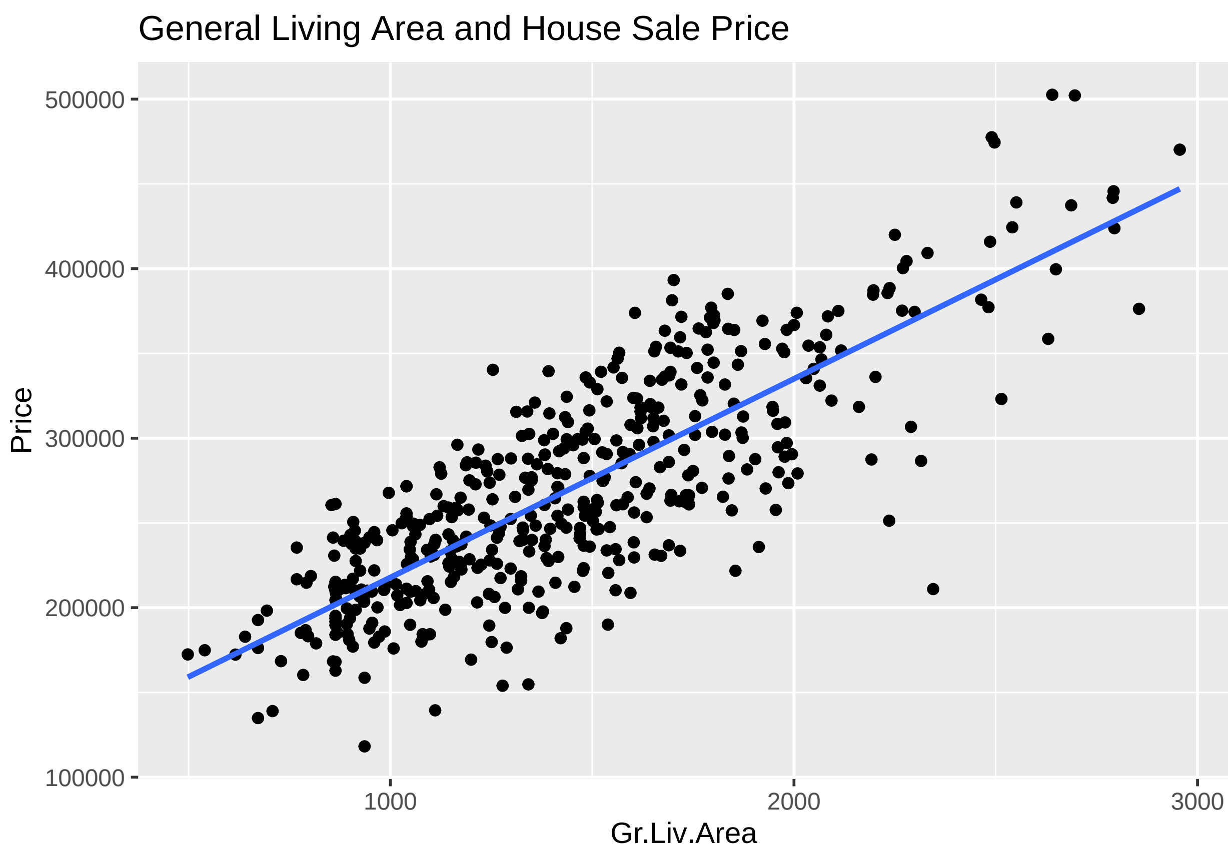


After fitting a simple linear regression model we can see that there are no obvious deviations from linearity that we need to account for. Additionally, we can see similar strong positive relationship between `Gr.Liv.Area` and `Price` and between `Garage.Cars` and `Price` as shown in the correlation plot above.

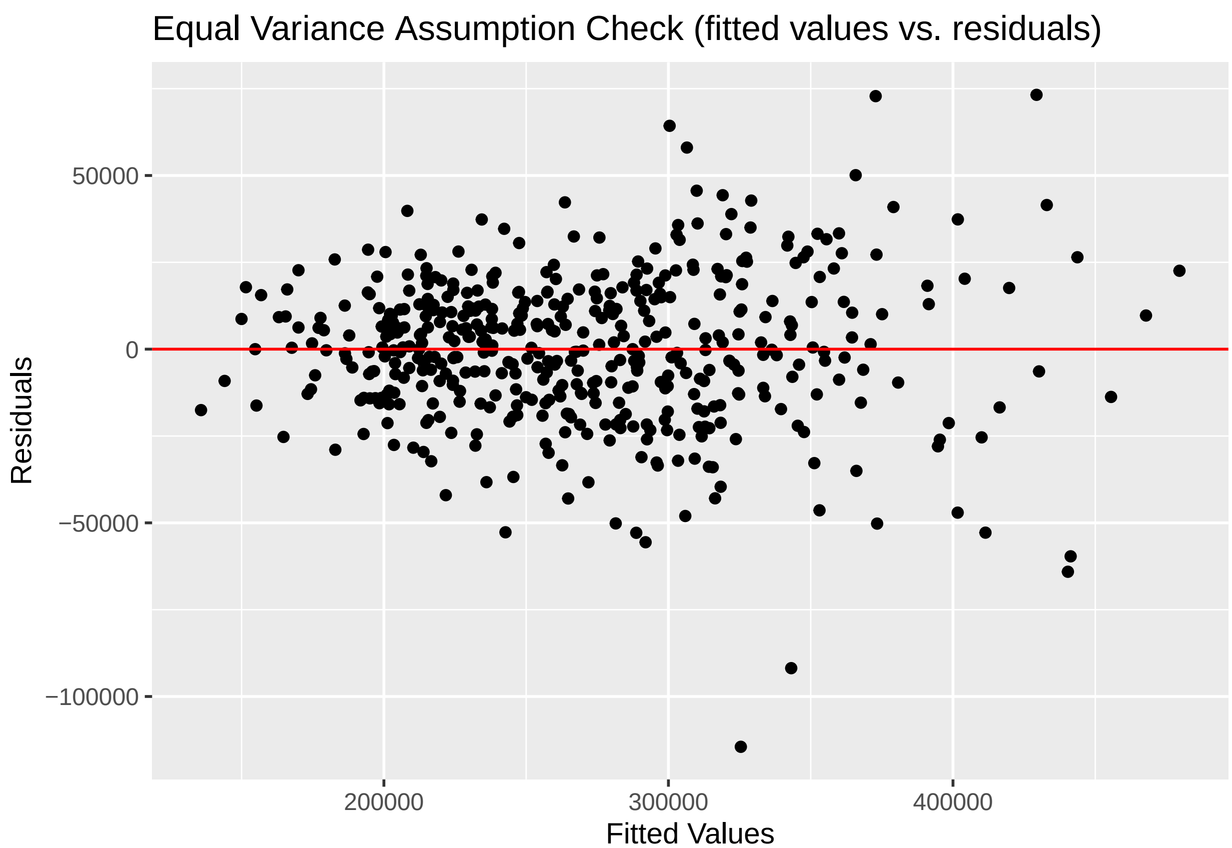

Price | others
Gr.Liv.Area | others
House.Style1.5Unf | others
House.Style1Story | others
House.Style2.5Fin | others
House.Style2.5Unf | others
House.Style2Story | others
House.StyleSFoyer | others
House.StyleSLvl | others
Year.Remod.Add | others



Finally, this plot seems to indicate that the residuals do not have equal variance and that the variance may increase as `Gr.Liv.Area` and `Price` increases (this will be shown later). The following plot of the residuals vs fitted values shows this problem of heteroskedasticity much more clearly. We may be able to fix some of this by a log transform of `Price`, but likely we should add a `weight` argument to a `gls()` model that accounts for the increased variability as `Gr.Liv.Area` increases.



General Living Area and House Sale Price
500000
400000
300000
200000
100000
Price
1000
2000
3000
Gr.Liv.Area



In summary, we must account for the spatial correlation of the residuals. If we ignore the spatial correlation this means that the residuals will not be independent of each other. Our predictions will still be unbiased, but our measure of standard error or precision are too small, resulting in artificially "significant" results. It also means that we are unable to use important information that is interdependence could otherwise provide us with. Finally, if we do not make the residuals homoskedastic we will still have unbiased estimates but wrong estimates of the standard errors.

Our proposed model is thus a general least squares model (`gls`) with `Price` as the target or explanatory variable. We will use the `weights` and `correlation` arguments to account for the heteroskedasticity and correlation respectively. We will set `weights = varExp(Gr.Liv.Area)` to account for the increase in variance as the size of the house increases. Finally, we will set `correlation = corExp(form=~Lon+Lat, nugget=TRUE)` to account for the spatial correlation explained by the `longitude` and `latitude` variables.

Section 2: Statistical Model

The model we chose for home appraisals is $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{DRD})$, with those variables expanded below:

where $\mathbf{y}$ is expanded to the matrix $\begin{bmatrix} \text{Price}_1 \\ \text{Price}_2 \\ \vdots \\ \text{Price}_{469} \end{bmatrix}$, where Price_i is the `Price` of a house appraisal at location i. Locations are determined by a transformed longitude and transformed latitude of a house's location (transformed so as to preserve anonymity). i represents house appraisals at locations 1, 2, . . . , 469.

where $\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,14} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,14} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{469,1} & x_{469,2} & \dots & x_{469,14} \end{bmatrix}$

where $x_{i,1}$ = The above ground living area in square feet of the house i

where $x_{i,2} = \begin{cases} 1 & \text{If the house style is "1.5Unf"} \\ 0 & \text{otherwise} \end{cases}$

where $x_{i,3} = \begin{cases} 1 & \text{If the house style is "1Story"} \\ 0 & \text{otherwise} \end{cases}$

where $x_{i,4} = \begin{cases} 1 & \text{If the house style is "2.5Fin"} \\ 0 & \text{otherwise} \end{cases}$

where $x_{i,5} = \begin{cases} 1 & \text{If the house style is "2.5Unf"} \\ 0 & \text{otherwise} \end{cases}$

where $x_{i,6} = \begin{cases} 1 & \text{If the house style is "2Story"} \\ 0 & \text{otherwise} \end{cases}$

where $x_{i,7} = \begin{cases} 1 & \text{If the house style is "SLvl"} \\ 0 & \text{otherwise} \end{cases}$

where $x_{i,8}$ = The remodel date (if no remodeling has been done, then the original construction date)

where $x_{i,9} = \begin{cases} 1 & \text{If the house has central air conditioning} \\ 0 & \text{otherwise} \end{cases}$

where $x_{i,10}$ = The number of full bathrooms above ground

where $x_{i,11}$ = The number of half bathrooms above ground

where $x_{i,12}$ = The number of bedrooms above ground

where $x_{i,13}$ = The number of kitchens above ground

where $x_{i,14}$ = The size of garage in car capacity

where $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_{\text{Gr.Liv.Area}} \\ \vdots \\ \beta_{\text{Garage.Cars}} \end{bmatrix}$

An example of how to interpret a numeric home characteristic can be shown with $\beta_{\text{Garage.Cars}}$, where, holding all other β coefficients constant, a 1 unit increase in $x_{i,14}$ (i.e., a garage holds 1 more car) results in a home appraisal price about $\beta_{\text{Garage.Cars}}$ higher, on average.

An example of how to interpret a categorical home characteristic can be shown with $\beta_{\text{House.Style1.5Unf}}$, where the difference of a home price appraisal for a 1.5 story home with an unfinished basement, compared to 1.5 story home with a finished basement, holding all else constant, results in a home appraisal price about $\beta_{\text{House.Style1.5Unf}}$ higher, on average.

where, for every unit increase in $x_{i,j}$, the price of a house i increases by β_j

where $\mathbf{D} = \begin{bmatrix} d_{1,1} & 0 & 0 & \dots & 0 \\ 0 & d_{2,2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_{469,469} \end{bmatrix}$. The $Var(y_i) = \sigma^2 d_{i,i}$ where σ^2 = The variance of all the residuals and $d_{i,i} = \exp(2\beta_{i,\text{Gr.Liv.Area}}\theta)$. Although θ is hard to interpret, if θ is positive, the variance increases as the values of $\beta_{\text{Gr.Liv.Area}}$ increase, but if θ is negative, the variance of the observations decreases as `Gr.Liv.Area` increases. We multiply the model by this matrix twice to account for the heteroskedasticity of `Gr.Liv.Area`.

Finally, the residuals of the model, or ϵ as written in most models, are determined by the correlation structure calculated by $\sigma^2\mathbf{R}(\phi, \omega)$). The matrix $\mathbf{R}$ is expanded to include

$$\begin{bmatrix} 1 & \rho(\text{Price}_1, \text{Price}_2) & \dots & \rho(\text{Price}_1, \text{Price}_{469}) \\ \rho(\text{Price}_2, \text{Price}_1) & 1 & \dots & \rho(\text{Price}_2, \text{Price}_{469}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(\text{Price}_{469}, \text{Price}_1) & \rho(\text{Price}_{469}, \text{Price}_2) & \dots & 1 \end{bmatrix}$$

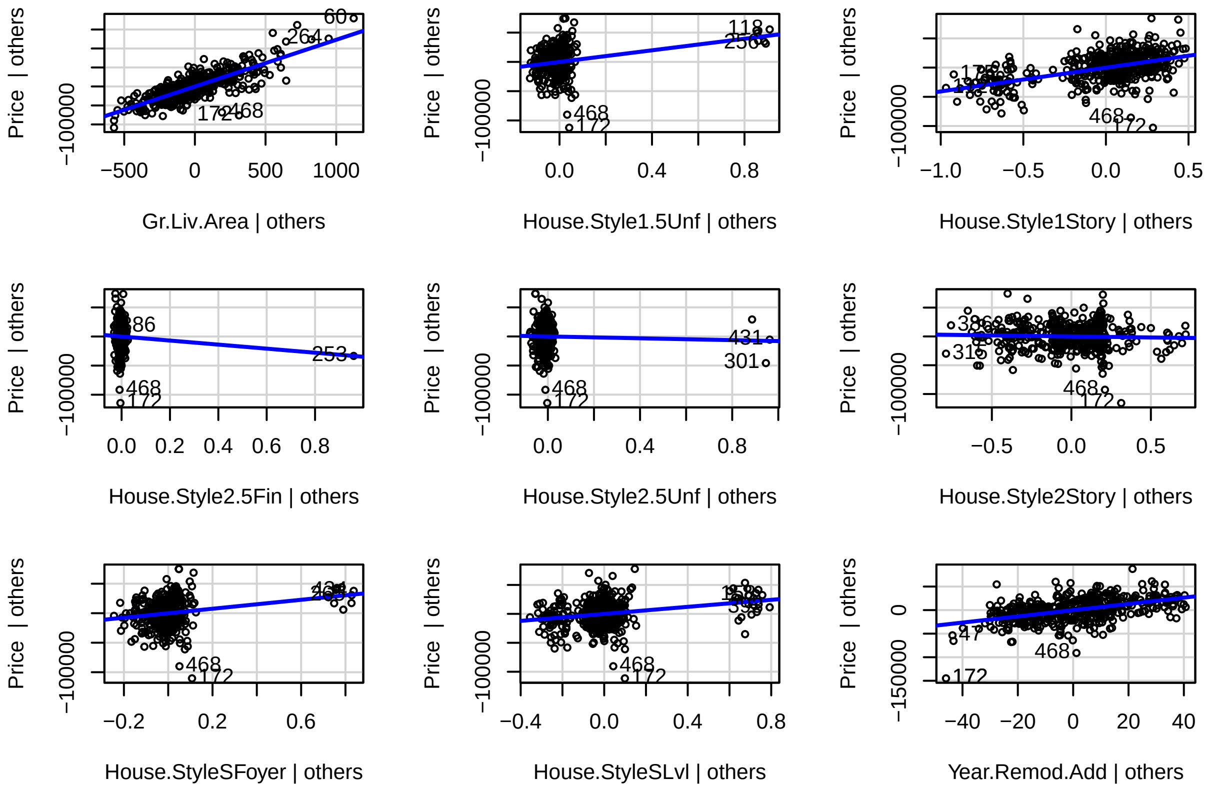
.

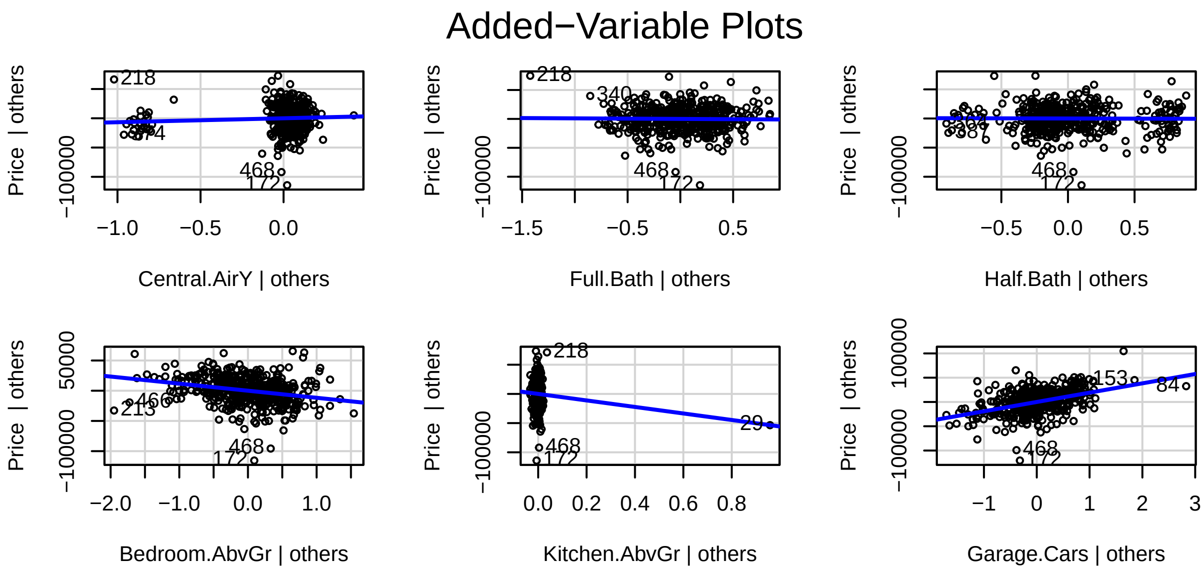
Specifically, this correlation is a general exponential correlation structure within `Price`, where $\rho(\text{Price}_i, \text{Price}_j) = -\exp\left\{\frac{\|\text{Price}_i - \text{Price}_j\|}{\phi}\right\}$, or the exponential correlation between locations i and j. The variables ϕ and ω are found through iterative optimization to help decorrelate the residuals in the $\mathbf{R}$ matrix, or $\rho(\text{Price}_i, \text{Price}_j)$ and use Maximum Likelihood Estimation to scale the variances over locations. The variable ϕ is the range of correlations between different locations. The variable ω is the correlation of a location with it's own self, and is used with $(1-\omega)\rho(\text{Price}_i, \text{Price}_j)$.

Assumptions: This model assumes that there is a linear relationship between $\mathbf{X}$ and $\mathbf{y}$. Additionally, we have accounted for the residuals interdependence by decorrelating the residuals by fitting a general exponential correlation structure on $\mathbf{y}$. We have accounted for the heteroskedasticity of the model by fitting diagonal matrices ($\mathbf{D}$) on the covariance matrix with the correlation matrix ($\mathbf{R}$). We now assume the model's resulting residuals are now independent with equal variance. Finally, we assume the residuals are normally distributed.

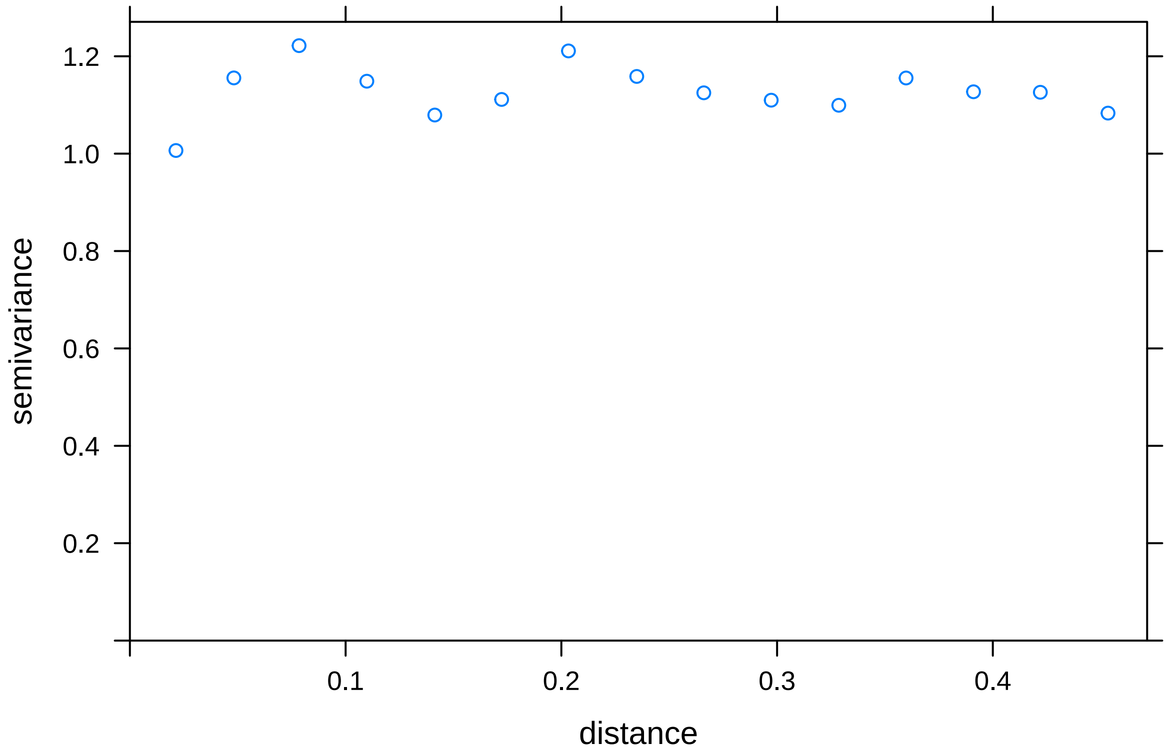
Section 3: Model Validation

After fitting the above model, we need to check that the assumptions are met.

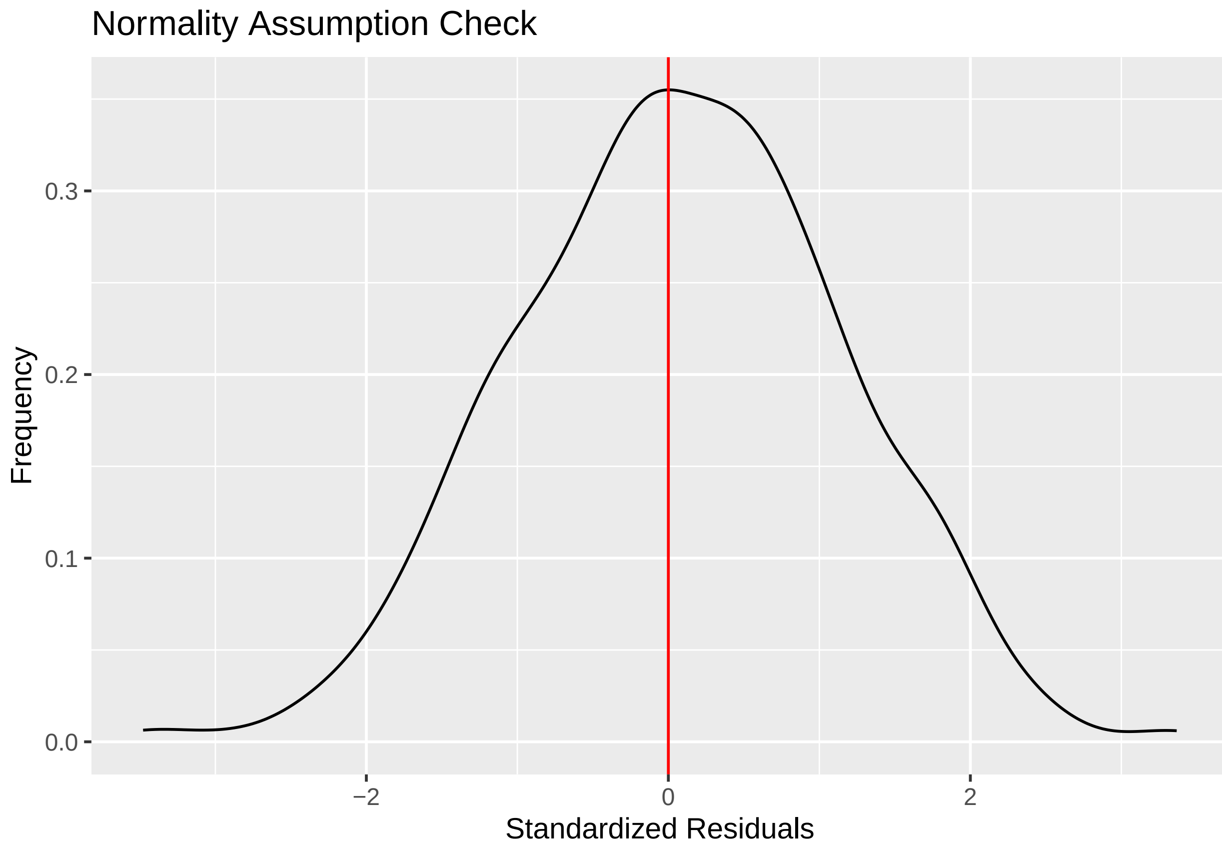

Price | others
Gr.Liv.Area | others
House.Style1.5Unf | others
House.Style1Story | others
House.Style2.5Fin | others
House.Style2.5Unf | others
House.Style2Story | others
House.StyleSFoyer | others
House.StyleSLvl | others
Year.Remod.Add | others



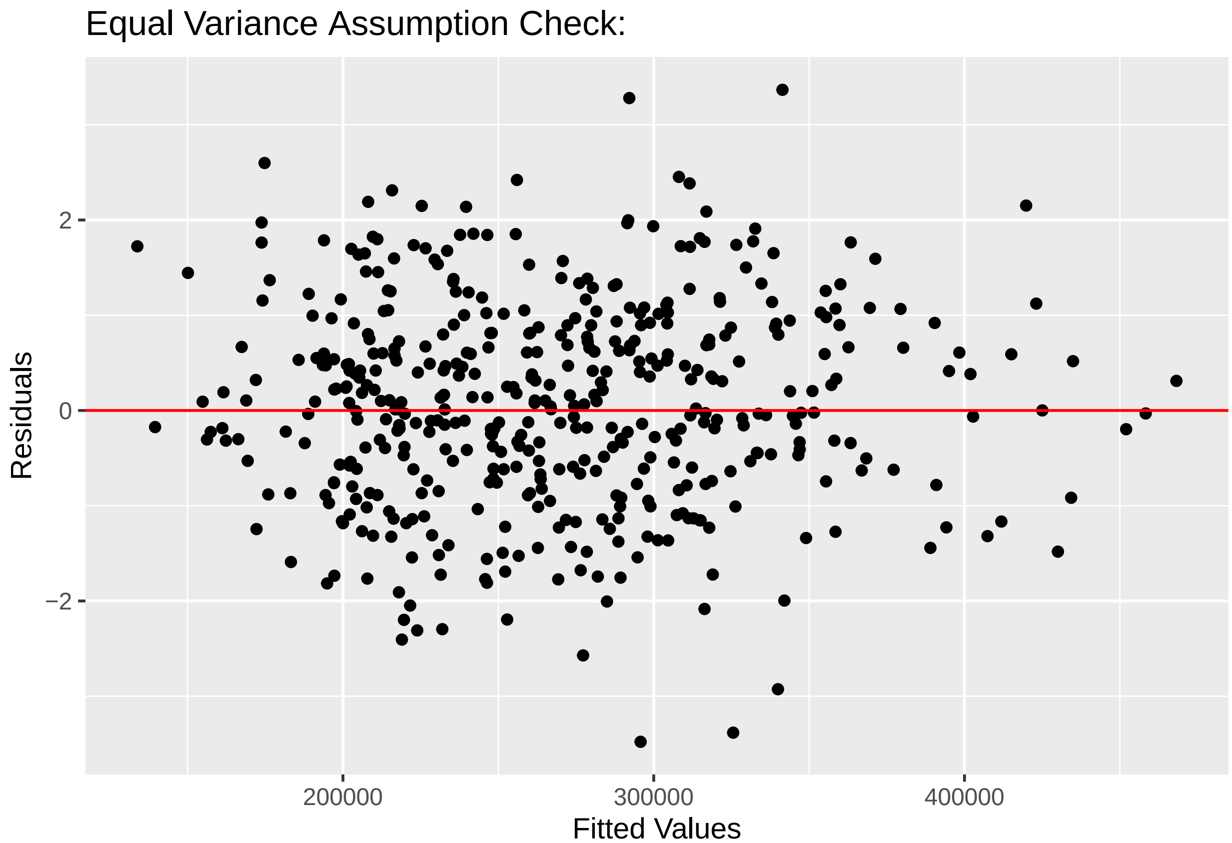
As before, these added-variable plots show that this assumption is met as there are no obvious deviations from linearity.



This variogram shows that the independence assumption is now met (forms a flat line) as the correlation introduced by location (`longitude` and `latitude`) is now accounted for in the model.



This graph of the residuals shows that they are fairly normally distributed and that the normality assumption is met.



This fitted values vs. residuals plot shows that the megaphone pattern is now gone and that the equal variance assumption is now met.

Additionally, the pseudo R^2 is 0.8892 which means that the model explains approximately 88.92% of the variance in house prices.

In order to validate predictions made with this model, we can perform a cross validation. After a 50 cross-validation studies the following metrics were as follows:
- Mean coverage = 0.96
- Mean bias = $-1506
- Mean width (of the confidence intervals) = $79704
- Mean RPMSE = $22386

The coverage of 0.96 indicates that it is correctly calculating the 95% confidence intervals such that it captures the true mean 95% of the time. The bias is not too large ($1506 is not very much compared to the average house price of $269547), although the width is a somewhat larger compared to the average house price. Finally, the RPMSE is much lower than the standard deviation of the house prices (65443) which is good as it means that the model helps to capture a portion of the variance of the house prices.

4. Section 4: Analysis Results To understand how well the home characteristics explain the home sale price we can refer to the pseudo R^2 calculated above which was 0.8892. This means that the home characteristics (the explanatory variables in the model) explain 88.92% of the variation in home sale price. In other words, the features provided in the data are very useful in predicting a home's sale price.

To determine what factors increase the sale price of a home, we can calculate 95% confidence intervals for the beta coefficients as shown below:

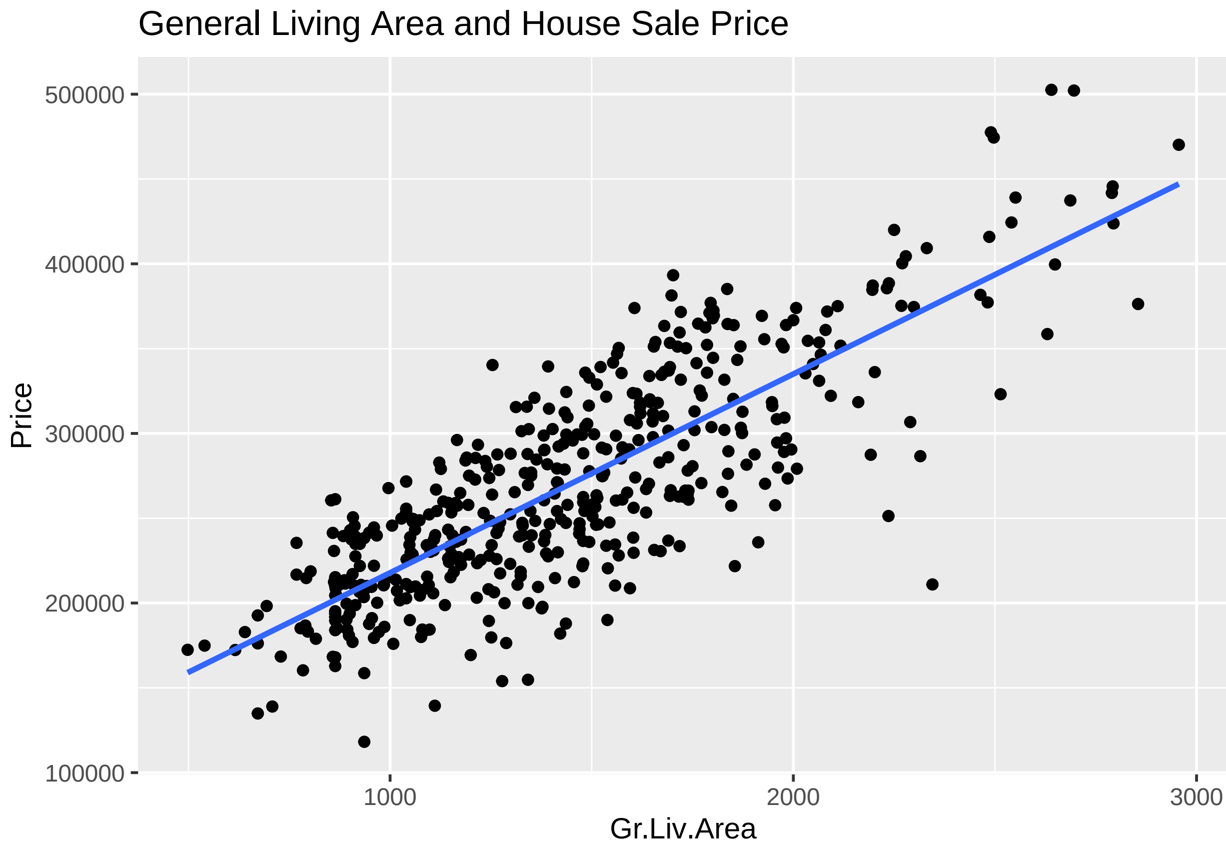
	2.5 %	97.5 %
(Intercept)	-1122177.4842	-897161.5954
Gr.Liv.Area	105.9819	123.1664
House.Style1.5Unf	32251.2228	47541.6914
House.Style1Story	32482.5886	39435.8589
House.Style2.5Fin	-100306.3338	-8839.0100
House.Style2.5Unf	-11684.4825	26719.7540
House.Style2Story	-5368.9256	4211.0949
House.StyleSFoyer	32734.8682	45690.8508
House.StyleSLvl	20812.1384	31910.8529
Year.Remod.Add	517.1452	627.7067
Central.AirY	8602.0432	16530.5596
Full.Bath	-2772.5244	4207.1817
Half.Bath	-2599.7336	3994.7402
Bedroom.AbvGr	-15217.8970	-11313.1039
Kitchen.AbvGr	-78591.3856	-29715.1557
Garage.Cars	19895.7104	23525.4549

Out of all of these features, those that significantly increase the house price in this model are:

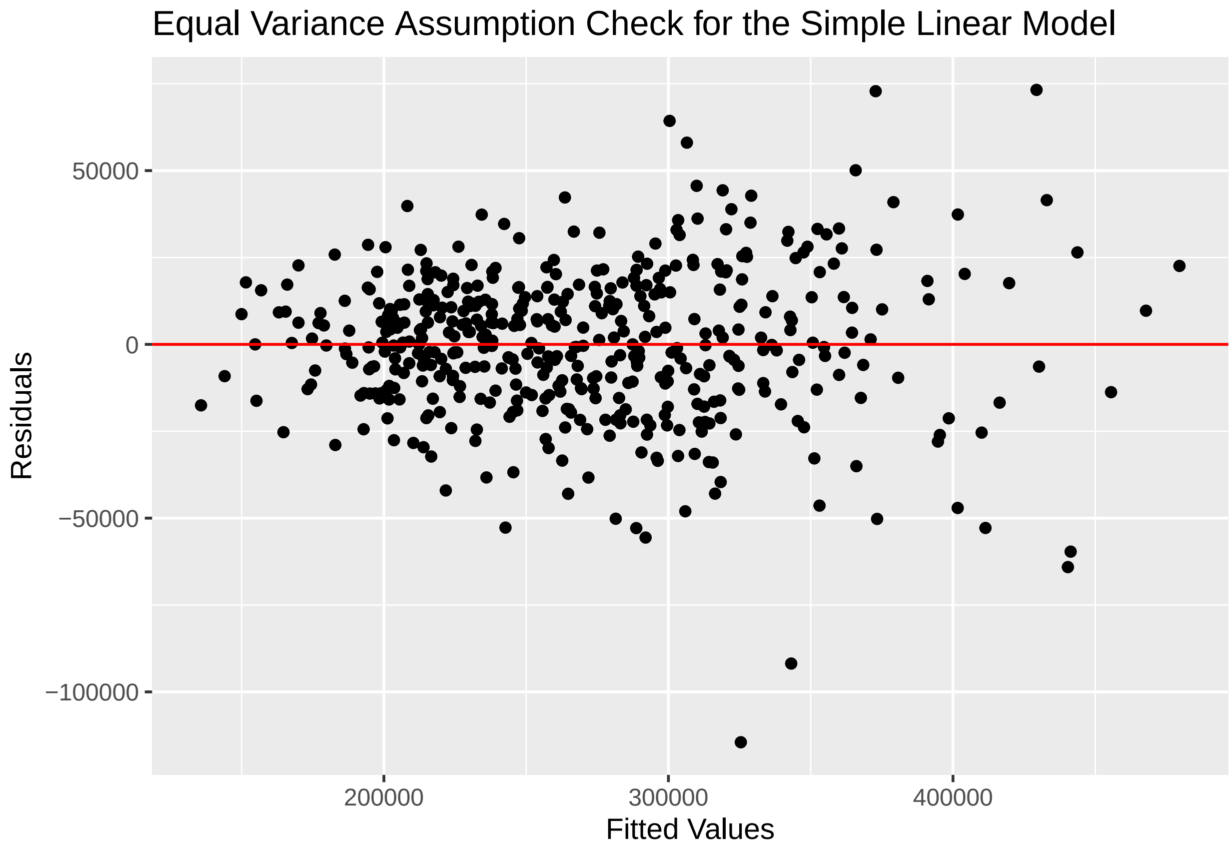
- The larger the general living area, the higher the price.
- The more recently the house has been remodeled, the higher the price.
- If there is central air, the higher the price.
- The larger the garage (the more cars that can fit into the garage), the higher the price
- Compared to having a `House.Style = "1.5Fin"`
  - Having a 1.5 level unfinished house (increases the price by at most $48000).
  - Having a 1 level house (increases the price by at most $39000).
  - Having an `"SFoyer"` style of house (increases the price by at most $46000).
  - Having an `"SLvl"` style of house (increases the price by at most $32000).

To determine if the variability of the sale price increases with the size of the home (as given by living area), we can do the following:

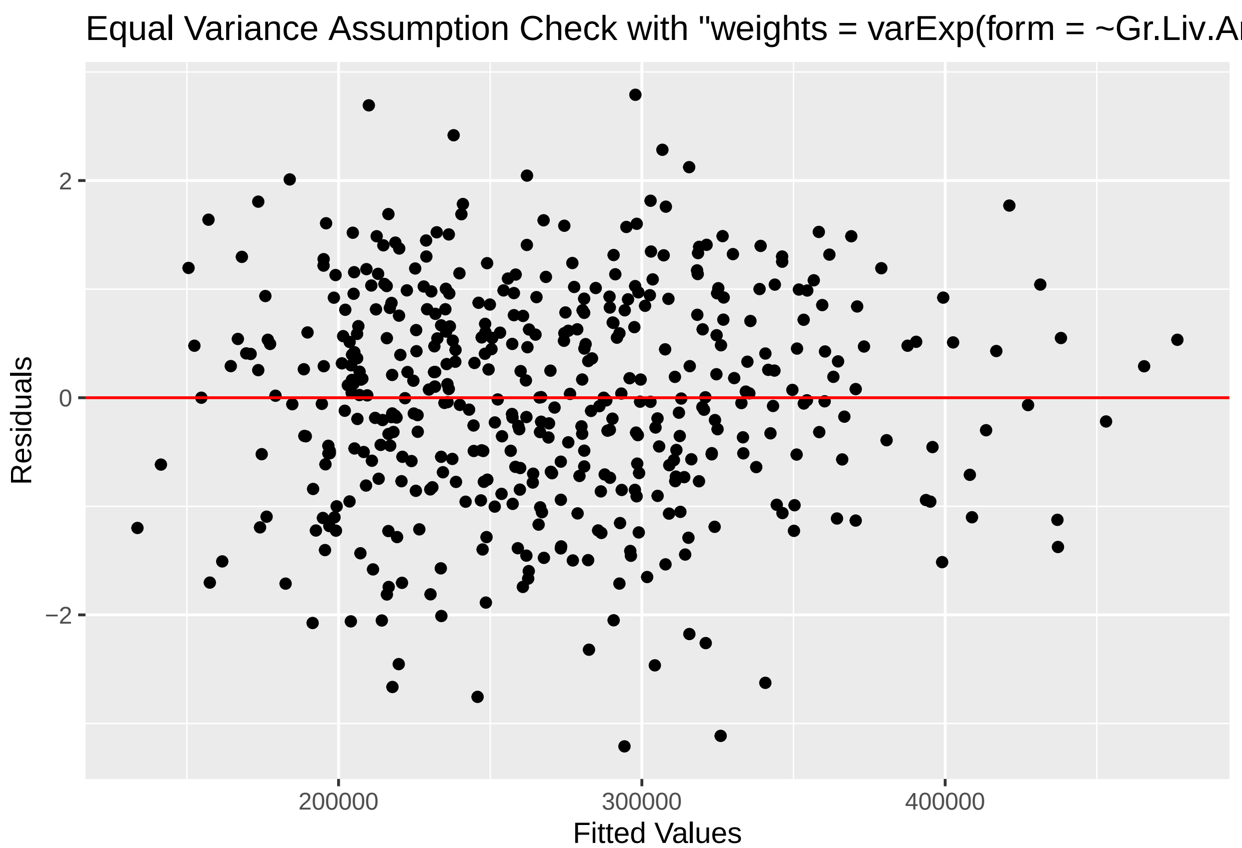
1. Plot the size of the home vs. the price of the home



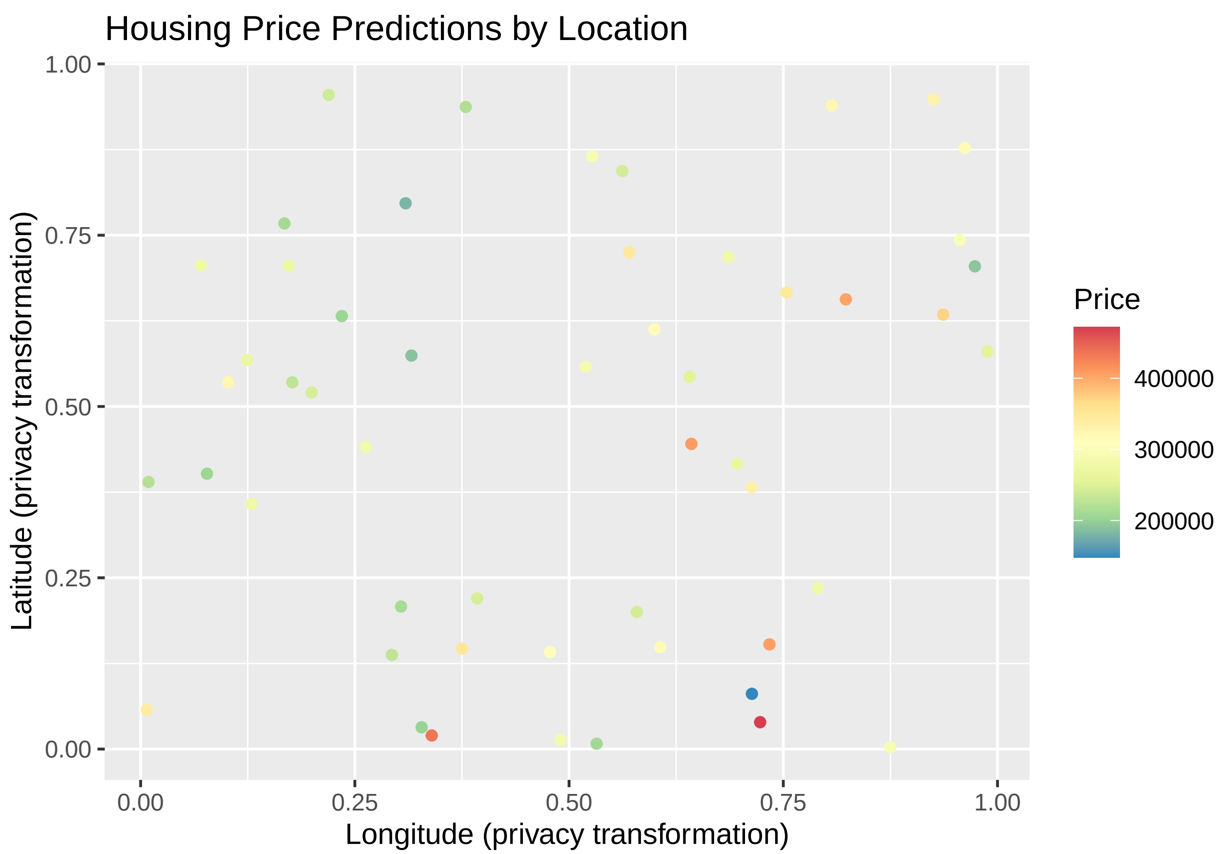
It is not immediately clear from this graph if the variance of the house sale price increases with its size. However, it is clear from this fitted values vs residuals plot that the variance of the sale price increases as sale price increases:



After fitting a `gls()` model with `weights = varExp(form = ~Gr.Liv.Area)` (meaning that the variance of the sale price increases as the size of the house increases), the resulting fitted values vs residuals plot no longer shows signs of heteroskedasticity as shown below. Because the heteroskedasticity went away when we added a function that described the variance of the house price increasing with its size, we believe that this is an indication that indeed the variability of the sale price increases with the size of the home.



Below is a map that depicts the predictions for the 52 locations with missing price values in the data set.



Section 5: Conclusions

From this analysis, we determined that the home characteristics explain home appraisal price fairly well, as evidenced by the fit measures we explored. The following factors, on average, lead to higher home appraisal prices: larger general living area, recent remodeling, central air, larger garage capacity, and having a style of house that was not 1.5 stories with a finished basement. We also determined that the home appraisal price increased with the general living area of a home. Finally, we predicted the home appraisal prices of 52 homes.

Some future concerns to keep in mind for more accurate home appraisal prices could include whether or not the house is part of an HOA and the state of the lawn or gardens surrounding the house. These factors could also determine the appraisal of a home's price.

Appendix of Code

```
knitr::opts_chunk$set(echo = FALSE, include = TRUE, fig.align = 'center')

# Load libraries
library(tidyverse) # for ggplot, dplyr, and magrittr
library(DataExplorer) # for correlation plot
library(GGally) # for ggpairs
library(car) # for checking linearity
library(gstat) # for variogram
```

```r
library(spdep) # for Moran and Geary tests
library(RColorBrewer) # to color graphs with more accessible color palette
library(nlme) # for GLM equation
library(gridExtra) # for plotting in groups
library(kableExtra) # for prettier report tables

# Source extra files
#source("../../../glstools-master/stdres.gls.R") # on Mary's machine
#source("../Drug Abuse/moranBasis.R") # on Mary's machine
source("~/R programming/STAT_469/moranBasis.R") # for Moran basis function # on Jillian's machine
source("~/R programming/STAT_469/stdres.gls.R") # for decorrelating residuals # on Jillian's machine
source("~/R programming/STAT_469/predictgls.R") # for predicting with gls # on Jillian's machine

# Set project options
options(scipen = 5) # for reducing scientific notation
set.seed(29) # for reproducibility

# Load data set
data <- read.csv("HousingPrices2.csv", stringsAsFactors = TRUE)
newdf <- data %>% filter(!is.na(Price)) # accounts for 52 missing prices
plot_correlation(newdf, type = "continuous")
ggplot(data = newdf, mapping = aes(x = Lon, y = Lat, color = Price)) +
  geom_point() +
scale_color_distiller(palette="Spectral", na.value = NA) +
    labs(title = "Housing Prices",
       x = "Longitude (privacy transformation)",
       y = "Latitude (privacy transformation)")
variogram(object=(Price)~Gr.Liv.Area
          +House.Style
          +Year.Remod.Add
          +Central.Air
          +Full.Bath
          +Half.Bath
          +Bedroom.AbvGr
          +Kitchen.AbvGr
          +Garage.Cars,
          locations=~Lon+Lat,
          data=newdf) %>%
  plot()
data_lm <- lm(formula = Price ~ .-Lon-Lat, data = newdf)
avPlots(data_lm, ask=FALSE)
newdf %>%
  ggplot(mapping = aes(x = Gr.Liv.Area, y = Price)) +
  geom_point() +
  geom_smooth(se = FALSE, method="lm") +
  labs(title="General Living Area and House Sale Price")
ggplot(mapping = aes(x=data_lm$fitted, y=data_lm$residuals)) +
  geom_point() +
  xlab('Fitted Values') +
  ylab('Residuals') +
  ggtitle('Equal Variance Assumption Check (fitted values vs. residuals)') +
  geom_hline(yintercept = 0, col = "red")
gls_model = gls(model=Price~Gr.Liv.Area
```

```
              +House.Style
              +Year.Remod.Add
              +Central.Air
              +Full.Bath
              +Half.Bath
              +Bedroom.AbvGr
              +Kitchen.AbvGr
              +Garage.Cars,
              data=newdf,
              correlation=corExp(form=~Lon+Lat, nugget=TRUE),
              method="ML",
              weights = varExp(form = ~Gr.Liv.Area))
# standardizing the residuals
sres = stdres.gls(gls_model)
data_lm <- lm(formula = Price ~ .-Lon-Lat, data = newdf)
avPlots(data_lm, ask=FALSE)
residDF <- data.frame(Lon=newdf$Lon, Lat=newdf$Lat, decorrResid=sres)
residVariogram <- variogram(object=decorrResid~1, locations=~Lon+Lat, data=residDF)
plot(residVariogram)
ggplot() +
  geom_density(mapping = aes(x = sres)) +
  xlab('Standardized Residuals') +
  ylab('Frequency') +
  ggtitle('Normality Assumption Check') +
  geom_vline(xintercept = 0, col = "red")
ggplot(mapping = aes(x=gls_model$fitted, y=sres)) +
  geom_point() +
  xlab('Fitted Values') +
  ylab('Residuals') +
  ggtitle('Equal Variance Assumption Check:') +
  geom_hline(yintercept = 0, col = "red")
# # Commented out so that it does not take forever to knit everytime.
# ## Run the CV code
# set.seed(59)
# n.cv <- 50 #Number of CV studies to run
#
# pb <- txtProgressBar(min = 0, max = n.cv, style = 3)
#
# n.test <- dim(data)[1]*0.2 #Number of observations in a test set
# rpmse <- rep(x=NA, times=n.cv)
# bias <- rep(x=NA, times=n.cv)
# wid <- rep(x=NA, times=n.cv)
# cvg <- rep(x=NA, times=n.cv)
#
# n = dim(data)[1]
#
# for(cv in 1:n.cv){
#   ## Select test observations
#   test.obs <- sample(x=1:n, size=n.test)
#
#   ## Split into test and training sets
#   test.set <- data[test.obs,]
#   train.set <- data[-test.obs,]
```

```
#
#   ## Fit a lm() using the training data
#   train.lm <- gls(model=Price~Gr.Liv.Area
#                 +House.Style
#                 +Year.Remod.Add
#                 +Central.Air
#                 +Full.Bath
#                 +Half.Bath
#                 +Bedroom.AbvGr
#                 +Kitchen.AbvGr
#                 +Garage.Cars,
#                 data=data,
#                 correlation=corExp(form=~Lon+Lat, nugget=TRUE),
#                 method="ML",
#                 weights = varExp(form = ~Gr.Liv.Area))
#
#   ## Generate predictions for the test set
#   my.preds <- predictgls(train.lm, newdframe = test.set)
#     # predict.lm(train.lm, newdata=test.set, interval="prediction")
#
#   ## Calculate bias
#   bias[cv] <- mean(my.preds[,'Prediction']-test.set[['Price']])
#
#   ## Calculate RPMSE
#   rpmse[cv] <- (test.set[['Price']]-my.preds[,'Prediction'])^2 %>% mean() %>% sqrt()
#
#   ## Calculate Coverage
#   cvg[cv] <- ((test.set[['Price']] > my.preds[,'lwr']) & (test.set[['Price']] < my.preds[,'upr'])) %>
#
#   ## Calculate Width
#   wid[cv] <- (my.preds[,'upr'] - my.preds[,'lwr']) %>% mean()
#
#   # Update the progress bar
#   setTxtProgressBar(pb, cv)
# }
#
# close(pb)
#
# cv_bias <- ggplot() +
#   geom_density(mapping = aes(x=bias)) +
#   xlab("Bias") +
#   geom_vline(xintercept = mean(bias), col = "red", lwd = 1)
#
# cv_rpmse <- ggplot() +
#   geom_density(mapping = aes(x=rpmse)) +
#   xlab("rpmse") +
#   ylab("Frequency") +
#   geom_vline(xintercept = mean(rpmse), col = "red", lwd = 1)
#
# cv_wid <- ggplot() +
#   geom_density(mapping = aes(x=wid)) +
#   xlab("wid") +
#   ylab("Frequency") +
```

```
#   geom_vline(xintercept = mean(wid), col = "red", lwd = 1)
#
# cv_cvg <- ggplot() +
#   geom_density(mapping = aes(x=cvg)) +
#   xlab("cvg") +
#   ylab("Frequency") +
#   geom_vline(xintercept = mean(cvg), col = "red", lwd = 1)
#
# grid.arrange(cv_rpmse, cv_bias, cv_wid, cv_cvg)
#
# paste("Mean coverage =",  mean(cvg))
# paste("Mean bias =", mean(bias))
# paste("Mean width =", mean(wid))
# paste("Mean RPMSE =", mean(rpmse))
confint(gls_model) %>%
  kable()
newdf %>%
  ggplot(mapping = aes(x=Gr.Liv.Area, y=Price)) +
  geom_point() +
  geom_smooth(se = F, method = "lm") +
  labs(title="General Living Area and House Sale Price")
ggplot(mapping = aes(x=data_lm$fitted, y=data_lm$residuals)) +
  geom_point() +
  xlab('Fitted Values') +
  ylab('Residuals') +
  ggtitle('Equal Variance Assumption Check for the Simple Linear Model') +
  geom_hline(yintercept = 0, col = "red")
gls_grlivarea <- gls(model=Price~Gr.Liv.Area
              +House.Style
              +Year.Remod.Add
              +Central.Air
              +Full.Bath
              +Half.Bath
              +Bedroom.AbvGr
              +Kitchen.AbvGr
              +Garage.Cars,
              data=newdf,
              method="ML",
              weights = varExp(form = ~Gr.Liv.Area))

sres_grlivarea = stdres.gls(gls_grlivarea)

ggplot(mapping = aes(x=gls_grlivarea$fitted, y=sres_grlivarea)) +
  geom_point() +
  xlab('Fitted Values') +
  ylab('Residuals') +
  ggtitle('Equal Variance Assumption Check with "weights = varExp(form = ~Gr.Liv.Area)"') +
  geom_hline(yintercept = 0, col = "red")

data_na <- read.csv("HousingPrices2.csv", stringsAsFactors = TRUE) %>%
  filter(is.na(Price))

data_pred = predictgls(gls_model, data_na) %>% #[c(1,4:12)]) %>%
```

```r
  mutate(Price = Prediction)

ggplot(data = data_pred, mapping = aes(x = Lon, y = Lat, color = Price)) +
  geom_point() +
  scale_color_distiller(palette="Spectral", na.value=NA) +
  labs(title = "Housing Price Predictions by Location",
       x = "Longitude (privacy transformation)",
       y = "Latitude (privacy transformation)")
#End of project's code
```