

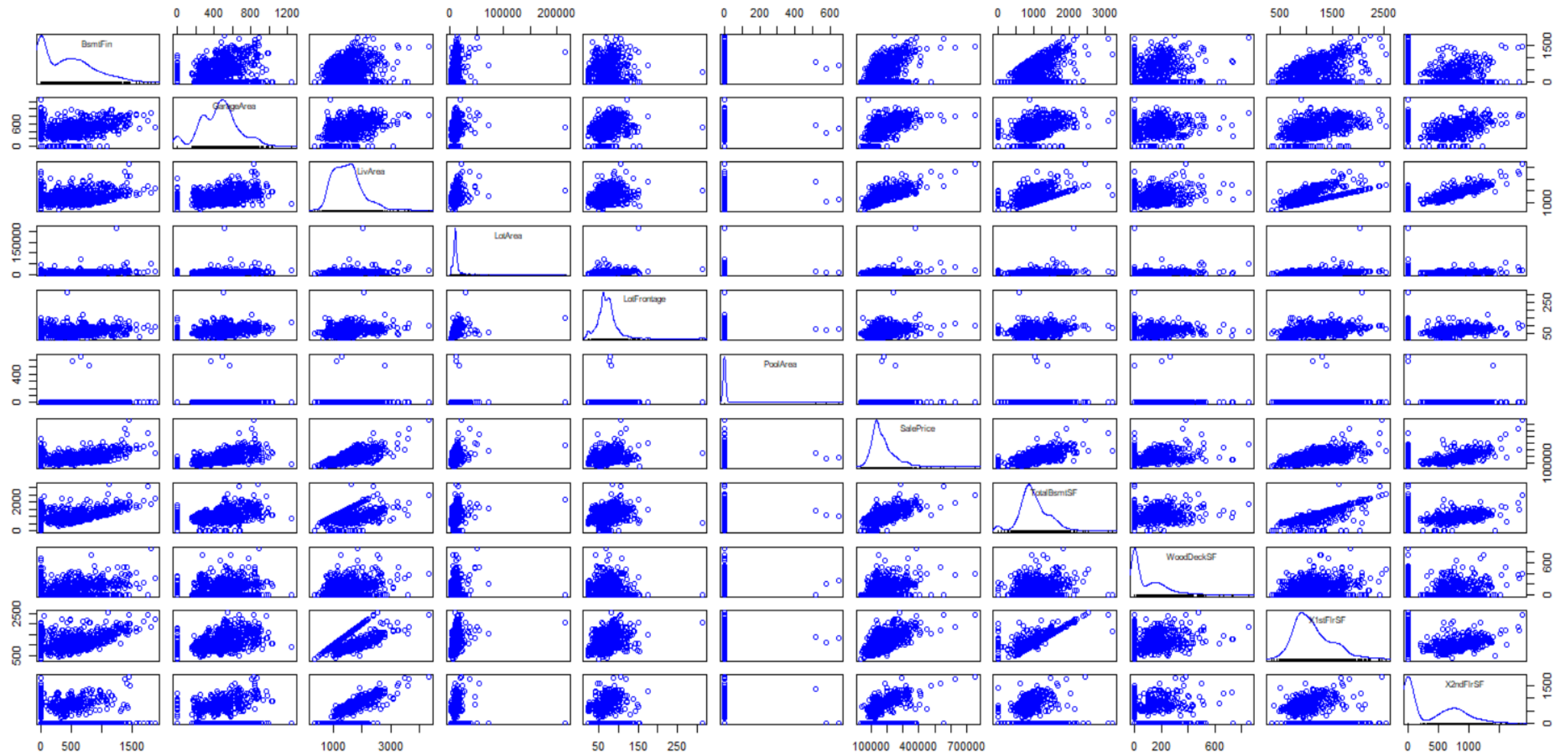
ML HW3

Variables

- Dependent Variable – Sale Price
- Independent Continuous Variables

1. LotFrontage
2. LotArea
3. BsmtFin
4. TotalBsmtSF
5. 1stFlrSF
6. 2ndFlrSF
7. LivArea
8. GarageArea
9. WoodDeckSF
10. PoolArea

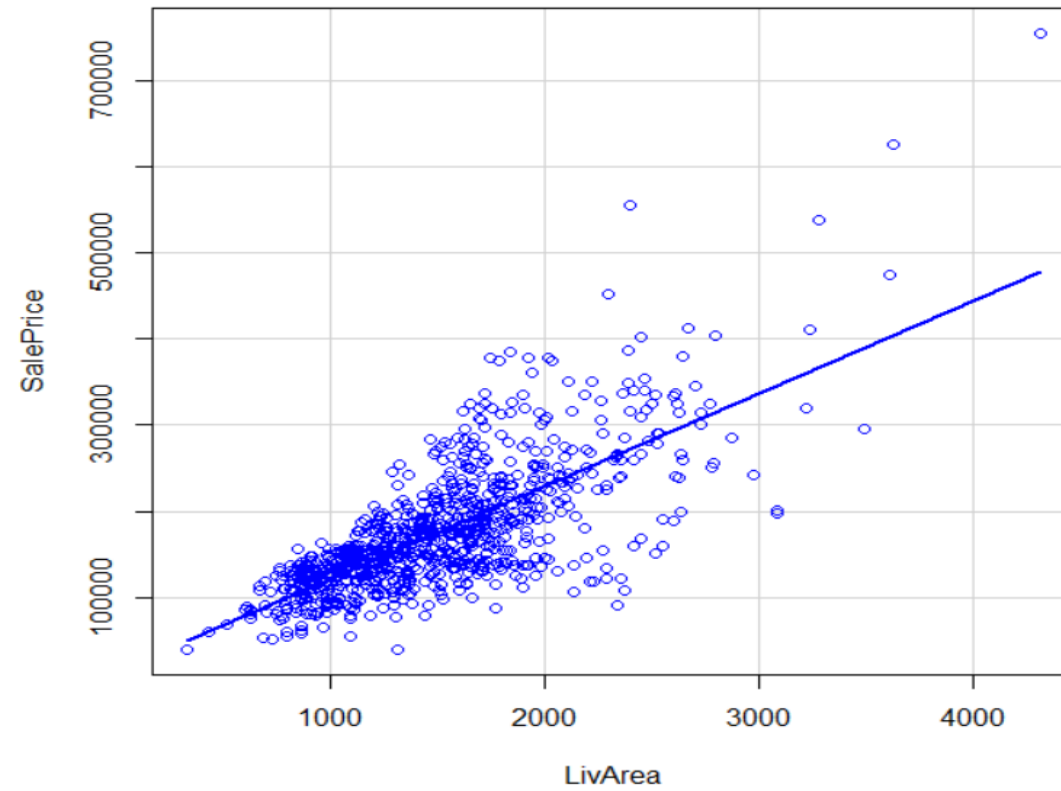
Scatterplot Matrix of Sales Price V/S All Continuous X Variables



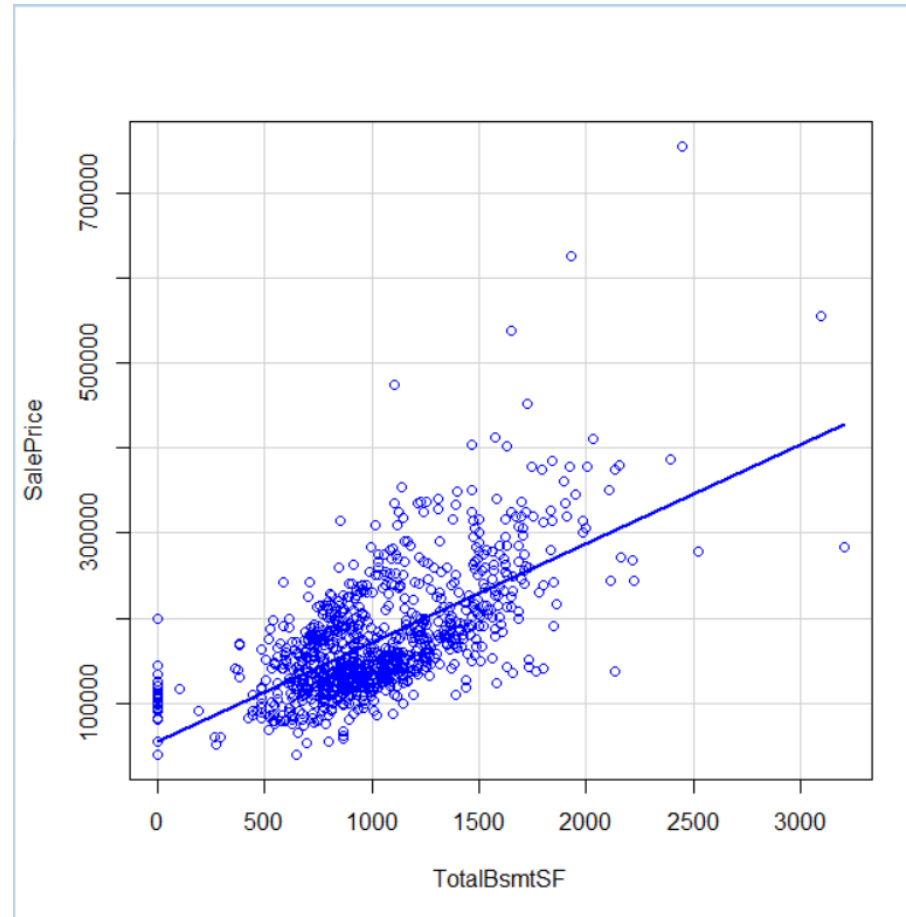
GRAPHS WITH STRONG RELATIONSHIPS BEWTEEN X-VARIABLE AND SALE PRICE

	<i>SalePrice</i>	<i>LotFrontage</i>	<i>LotArea</i>	<i>BsmtFin</i>	<i>TotalBsmtSF</i>	<i>1stFlrSF</i>	<i>2ndFlrSF</i>	<i>LivArea</i>	<i>GarageArea</i>	<i>WoodDeckSF</i>	<i>PoolArea</i>
SalePrice	1										
LotFrontage	0.343362674	1									
LotArea	0.345210323	0.367733798	1								
BsmtFin	0.429429111	0.153806408	0.183688798	1							
TotalBsmtSF	0.648682195	0.290481555	0.251400128	0.470190323	1						
1stFlrSF	0.624208452	0.395408286	0.295664287	0.39061046	0.789542893	1					
2ndFlrSF	0.347301593	0.047042928	0.050817554	-0.150695214	-0.17942366	-0.21229	1				
LivArea	0.738919278	0.31999123	0.253237165	0.137863664	0.39901765	0.522782	0.716377	1			
GarageArea	0.616705484	0.298245784	0.18902382	0.263659319	0.447885356	0.453221	0.153807	0.447201	1		
WoodDeckSF	0.37909777	0.072435969	0.137573318	0.22511471	0.241763046	0.250303	0.11583	0.274614	0.235516734	1	
PoolArea	0.019033392	0.021363044	0.013860875	0.032505392	0.019805616	0.022966	0.007999	0.022386	0.002818642	0.033478223	1

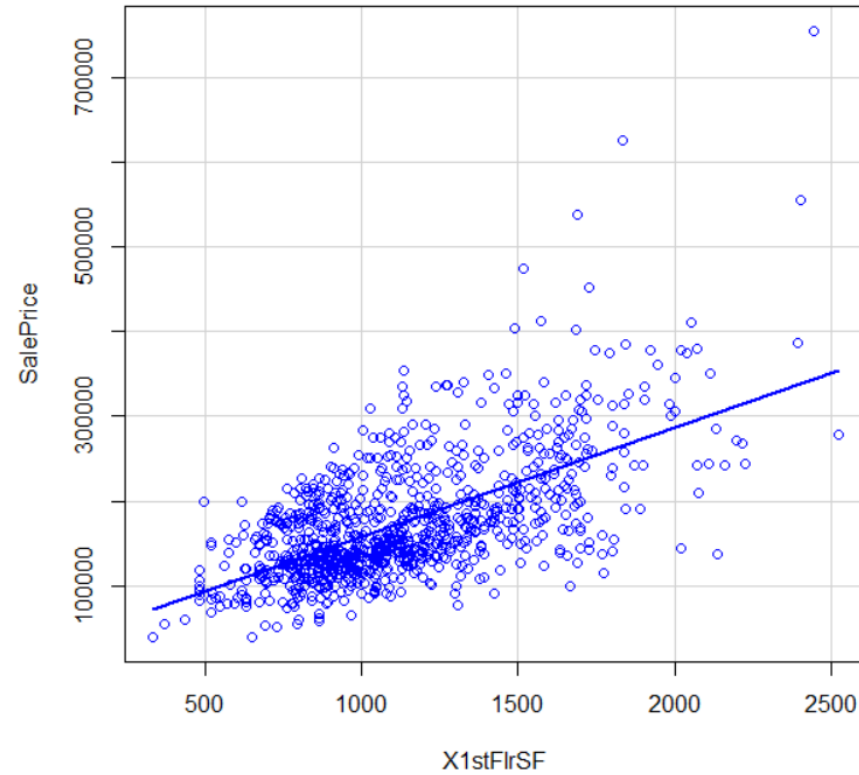
LivArea and SalesPrice



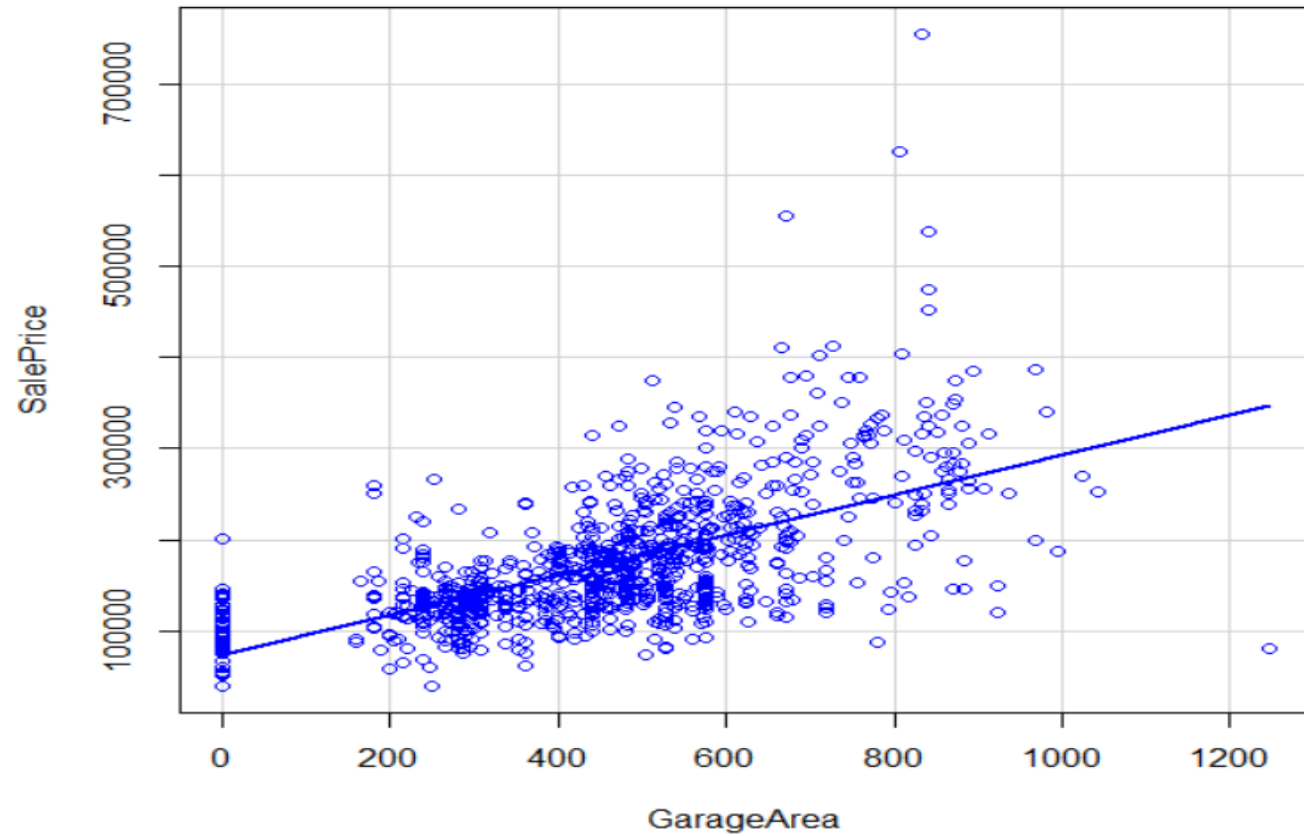
TotalBsmtSF and SalesPrice



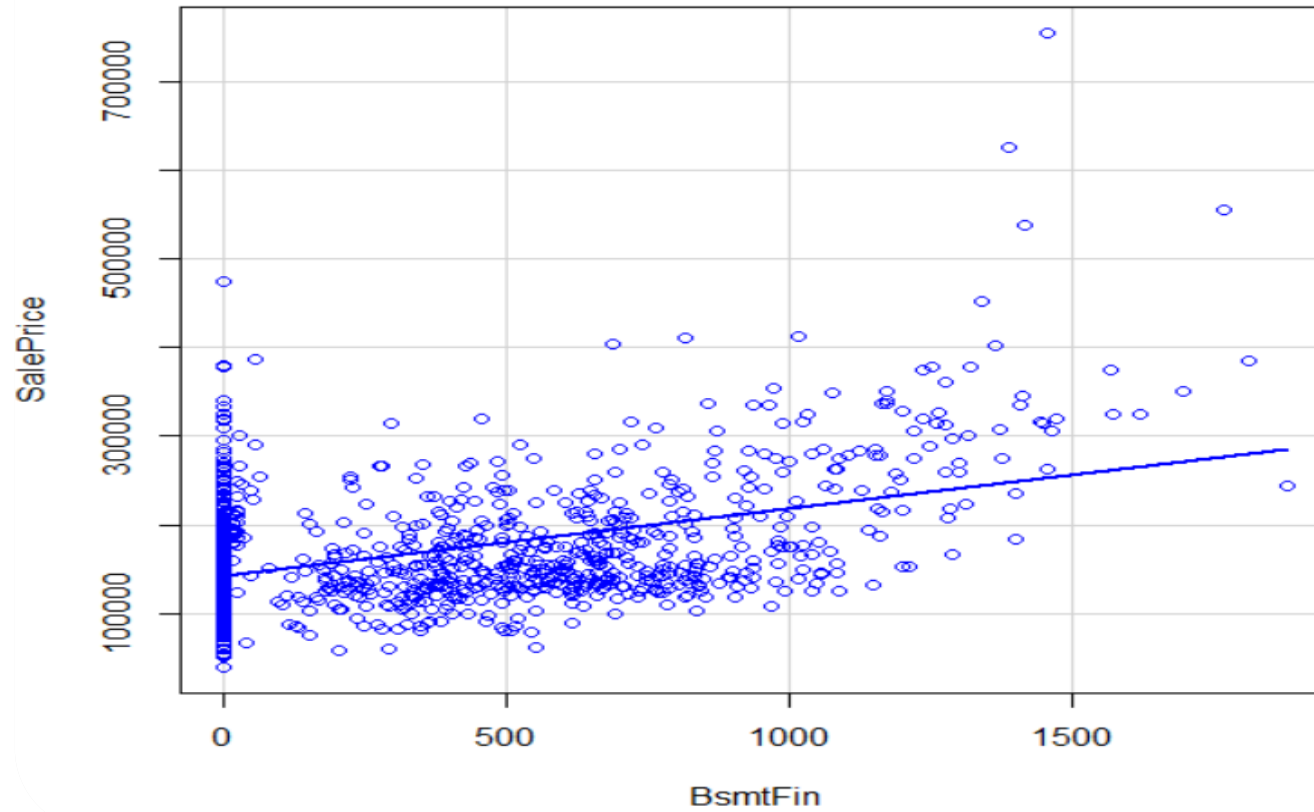
1stFlrSF and Sales Price



GarageArea and Sales Price



BsmtFin and Sales Price



VIF

```
> vif(RegModel.2)
```

Bedroom	BsmtFin	BsmtFullBath	BsmtHalfBath	Fireplaces	FullBath	GarageArea	GarageCars	HalfBath	Kitchen	LivArea	LotArea	LotFrontage	OverallCond
2.228450	2.237026	1.933482	1.128702	1.499121	2.805047	5.110263	5.505599	2.177244	1.383041	111.624502	1.253143	1.424288	1.364292
OverallQual	PoolArea	TotalBsmtSF	TotRmsAbvGrd	WoodDeckSF	X1stFlrSF	X2ndFlrSF	YearBuilt						
2.924626	1.015221	3.402420	4.729238	1.198643	58.777007	83.331675	3.002037						

VIF Analysis

ATTRIBUTE	VIF
PoolArea	1.015221
BsmtHalfBath	1.128702
WoodDeckSF	1.198643
LotArea	1.253143
OverallCond	1.364292
Kitchen	1.383041
LotFrontage	1.424288
Fireplaces	1.499121
BsmtFullBath	1.933482
HalfBath	2.177244
Bedroom	2.22845
BsmtFin	2.237026
FullBath	2.805047
OverallQual	2.924626
YearBuilt	3.002037
TotalBsmtSF	3.40242
TotRmsAbvGrd	4.729238
GarageArea	5.110263
GarageCars	5.505599
X1stFlrSF	58.777007
X2ndFlrSF	83.331675
LivArea	111.624502

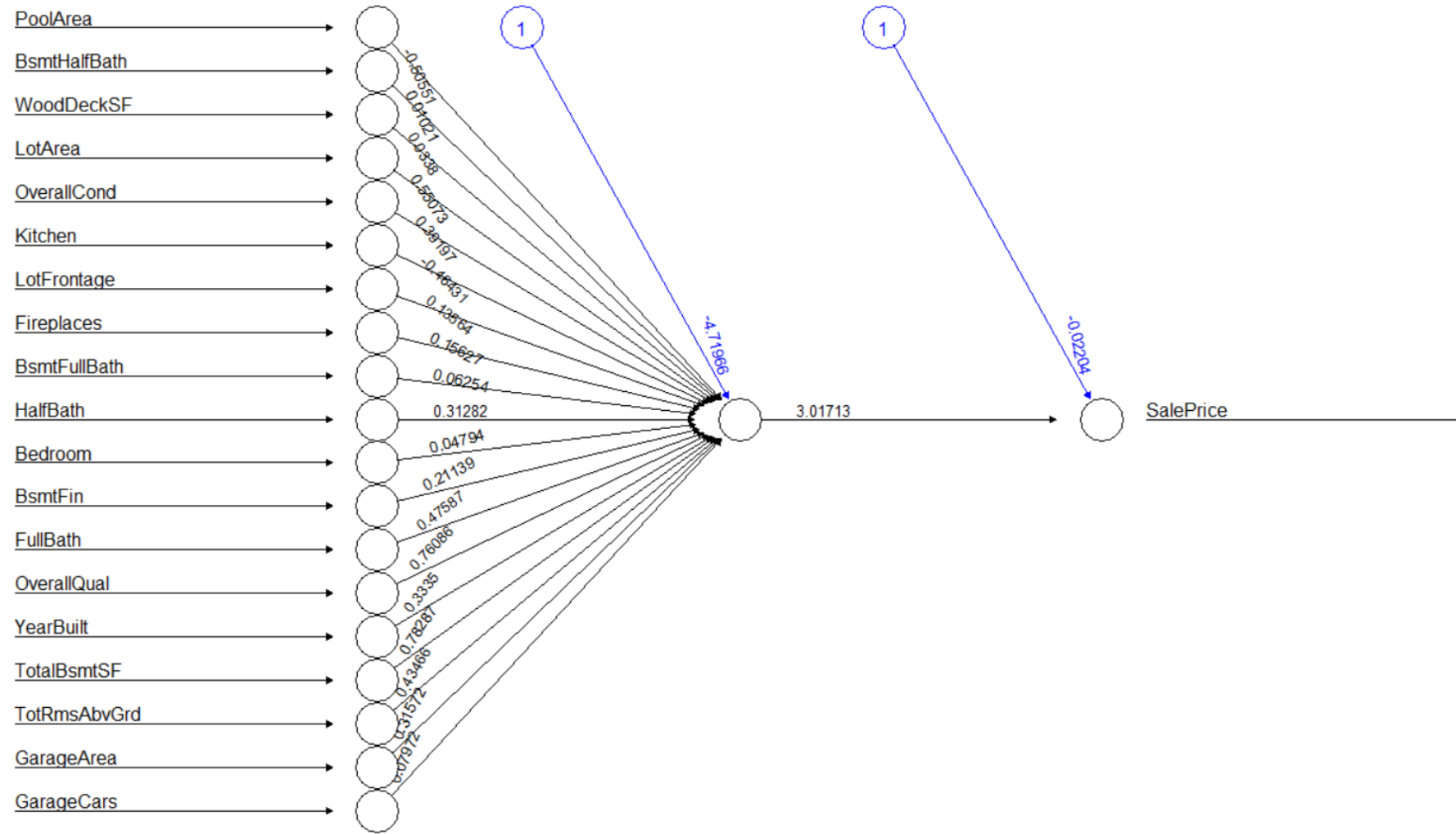
VIF <10

Attribute	VIF
PoolArea	1.015221
BsmtHalfBath	1.128702
WoodDeckSF	1.198643
LotArea	1.253143
OverallCond	1.364292
Kitchen	1.383041
LotFrontage	1.424288
Fireplaces	1.499121
BsmtFullBath	1.933482
HalfBath	2.177244
Bedroom	2.22845
BsmtFin	2.237026
FullBath	2.805047
OverallQual	2.924626
YearBuilt	3.002037
TotalBsmtSF	3.40242
TotRmsAbvGrd	4.729238
GarageArea	5.110263
GarageCars	5.505599

NN WITH 1 HIDDEN LAYER

- `install.packages("neuralnet", dependencies = TRUE)`
- `library(neuralnet)`
- `housing_index <- sample(nrow(Housing), 1/2 * nrow(Housing))`
- `housing_train <- Housing[housing_index,]`
- `housing_test <- Housing[-housing_index,]`
- `head(housing_train)`
- `head(housing_test)`
- `normalize <- function(x) {return((x-min(x))/(max(x)-min(x)))}`
- `trainingnorm <- as.data.frame(lapply(housing_train,normalize))`
- `testingnorm <- as.data.frame(lapply(housing_test,normalize))`
- `housingnet <- neuralnet(SalePrice ~ PoolArea+BsmtHalfBath+ WoodDeckSF+ LotArea+ OverallCond+ Kitchen+ LotFrontage+ Fireplaces+ BsmtFullBath+ HalfBath+ Bedroom+ BsmtFin+ FullBath+ OverallQual+ YearBuilt+ TotalBsmtSF+ TotRmsAbvGrd+ GarageArea+ GarageCars`
- `, trainingnorm, lifesign="minimal", linear.output=TRUE, threshold=0.01)`
- `plot(housingnet)`

Housenet Plot



Testing Data

- `temp_test <- subset(testingnorm, select = c("PoolArea","BsmtHalfBath","WoodDeckSF",`
- `"LotArea","OverallCond","Kitchen","LotFrontage","Fireplaces","BsmtFullBath",`
- `"HalfBath","Bedroom","BsmtFin","FullBath", "OverallQual", "YearBuilt", "TotalBsmtSF", "TotRmsAbvGrd",`
`"GarageArea", "GarageCars"))`
- `head(temp_test)`
- `housingnet_results <- compute(housingnet, temp_test)`
- `predicted_price <- housingnet_results$net.result`
- `cor(predicted_price, testingnorm$SalePrice)`

Testing Data

```
> head(temp_test)
```

	PoolArea	BsmtHalfBath	WoodDeckSF	LotArea	OverallCond	Kitchen	LotFrontage	Fireplaces	BsmtFullBath	HalfBath	Bedroom	BsmtFin	FullBath	OverallQual	YearBuilt	TotalBsmtSF
1	0	1	0.34772462	0.1552212	0.8571429	0	0.4503817	0.3333333	0.0	0.0	0.5000000	0.5403315	0.6666667	0.5555556	0.7591241	0.3936369
2	0	0	0.04667445	0.2420966	0.4285714	0	0.4885496	0.0000000	0.5	0.5	0.1666667	0.4044199	0.3333333	0.4444444	0.8832117	0.2482845
3	0	0	0.00000000	0.1132747	0.5714286	1	0.2213740	0.6666667	0.5	0.0	0.3333333	0.4701657	0.3333333	0.4444444	0.4890511	0.3091079
4	0	0	0.00000000	0.1781378	0.4285714	1	0.3893130	0.0000000	0.0	0.0	0.3333333	0.0000000	0.6666667	0.3333333	0.6934307	0.0000000
5	0	0	0.00000000	0.2340151	0.4285714	0	0.3435115	0.0000000	0.5	0.5	0.5000000	0.3569061	0.3333333	0.4444444	0.9635036	0.3474735
6	0	0	0.19953326	0.1579535	0.4285714	0	0.4122137	0.3333333	0.0	0.0	0.5000000	0.0000000	0.6666667	0.7777778	0.9489051	0.5542732

	TotRmsAbvGrd	GarageArea	GarageCars
1	0.4	0.4410355	0.6666667
2	0.3	0.4602109	0.6666667
3	0.3	0.1965484	0.3333333
4	0.4	0.4947267	0.6666667
5	0.4	0.5522531	0.6666667
6	0.5	0.5119847	0.6666667

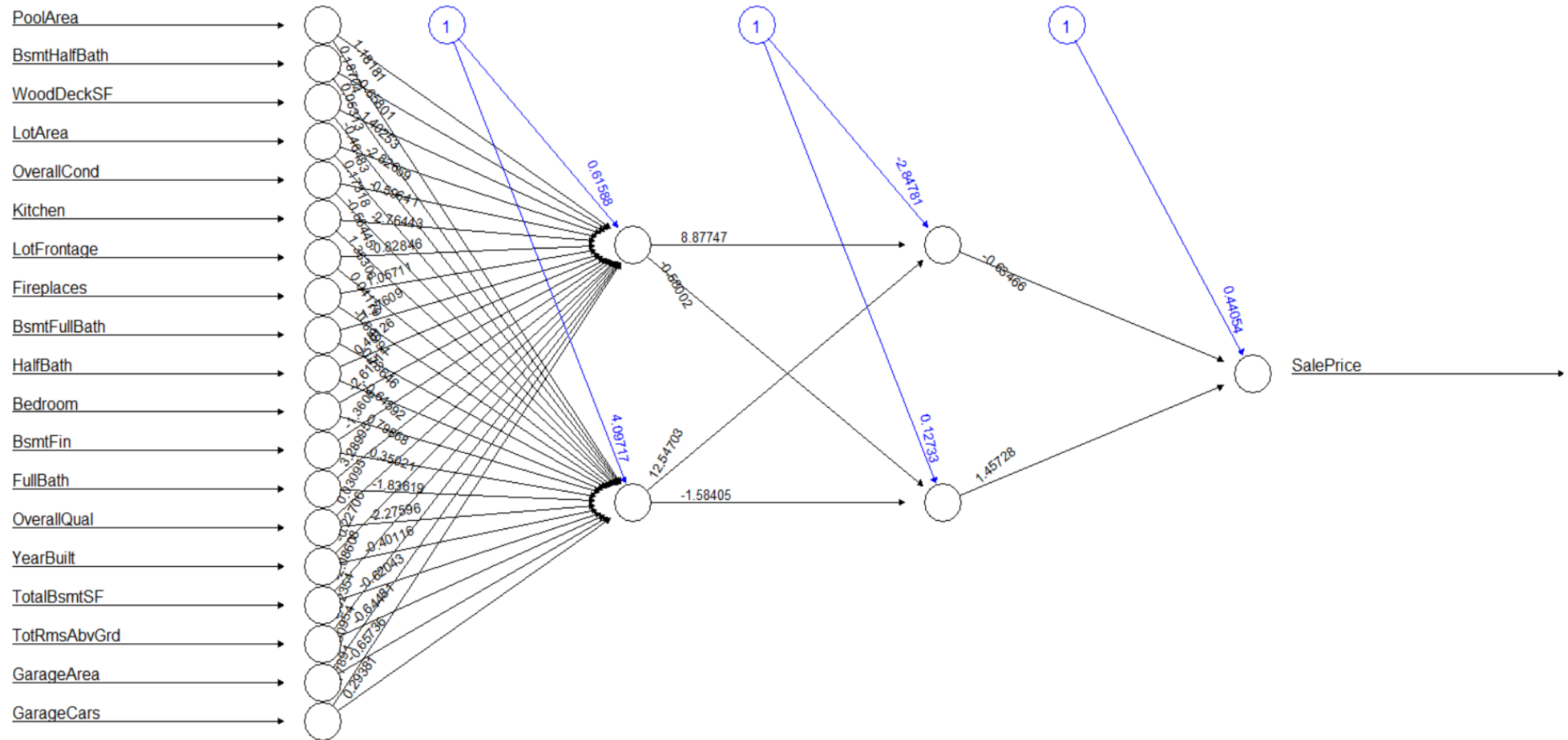
Correlation between Predicted Price & Actual Price

```
> cor(predicted_price, testingnorm$SalePrice)
      [,1]
[1,] 0.9214179
> |
```

NN WITH 2 HIDDEN LAYERS

- `install.packages("neuralnet", dependencies = TRUE)`
- `library(neuralnet)`
- `housing_index <- sample(nrow(Housing), 1/2 * nrow(Housing))`
- `housing_train <- Housing[housing_index,]`
- `housing_test <- Housing[-housing_index,]`
- `head(housing_train)`
- `head(housing_test)`
- `normalize <- function(x) {return((x-min(x))/(max(x)-min(x)))}`
- `trainingnorm <- as.data.frame(lapply(housing_train,normalize))`
- `testingnorm <- as.data.frame(lapply(housing_test,normalize))`
- `housingnet <- neuralnet(SalePrice ~PoolArea+BsmtHalfBath+WoodDeckSF+LotArea+OverallCond`
- `+ Kitchen+LotFrontage+Fireplaces+BsmtFullBath+HalfBath+Bedroom+ BsmtFin+ FullBath`
- `+OverallQual+YearBuilt+TotalBsmtSF+TotRmsAbvGrd+ GarageArea+ GarageCars, trainingnorm, hidden=c(2,2), lifesign="minimal", linear.output=TRUE, threshold=0.01)`
- `plot(housingnet)`

NN 2 Plot



Prediction on 50% Random Testing Data

- `temp_test <- subset(testingnorm, select = c("PoolArea","BsmtHalfBath","WoodDeckSF",`
- `"LotArea","OverallCond","Kitchen","LotFrontage","Fireplaces","BsmtFullBath",`
- `"HalfBath","Bedroom","BsmtFin","FullBath", "OverallQual", "YearBuilt", "TotalBsmtSF", "TotRmsAbvGrd",`
`"GarageArea", "GarageCars"))`
- `head(temp_test)`
- `housingnet_results <- compute(housingnet, temp_test)`
- `predicted_price <- housingnet_results$net.result`
- `cor(predicted_price, testingnorm$SalePrice)`

```
> cor(predicted_price, testingnorm$SalePrice)
      [,1]
[1,] 0.941191
> |
```

COMPARISON BETWEEN DIFFERENT NN

CORRELATION	1 Hidden Layer, 1 Node	1 Hidden Layer, 2 Nodes	1 Hidden Layer, 3 Nodes	2 Hidden Layers - c (1,1)	2 Hidden Layers - c(1,2)	2 Hidden Layers -c(2,2)	2Hidden Layers-c(3,3)	2Hidden Layers-c(4,4)
set.seed(1)	0.95	0.94	0.95	0.94	0.09	0.95	0.95	0.94
set.seed(2)	0.94	0.92	0.93	-0.16	0.09	0.94	0.93	0.85
set.seed(3)	-0.13	0.91	0.9	-0.12	0.95	0.89	0.88	0.9
set.seed(5)	0.93	0.92	0.92	0.93	0.93	0.92	0.91	0.9
set.seed(6)	0.94	0.94	0.9	0.94	0.94	0.92	0.93	0.91
set.seed(7)	0.94	0.92	0.92	0.92	0.92	0.92	0.92	0.92
set.seed(8)	0.93	0.92	0.92	0.93	0.94	0.87	0.93	0.85
set.seed(10)	0.94	0.92	0.92	0.94	0.94	0.87	0.91	0.89
Average Correlation	0.81	0.92	0.92	0.67	0.73	0.91	0.92	0.90