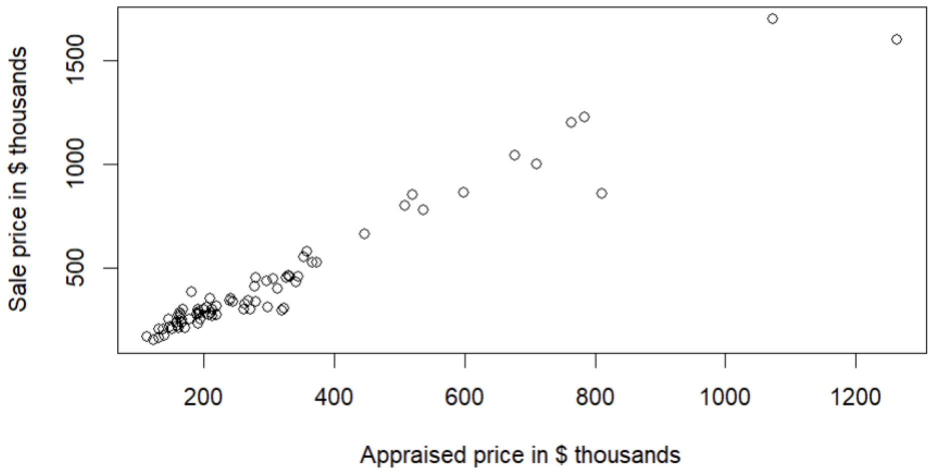
Activity 3_1:

(a) Propose a straight-line model to relate the appraised property value *x* to the sale price *y* for residential properties in this neighborhood.

From the scatterplot we observe linear relationship between the two variables.




```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.35868   13.76817   0.099    0.922
## xvariable    1.40827    0.03693  38.132   <2e-16 ***
```

Predicted_sale_price = 1.35868 + appraised_value * 1.40827
* Both the predicted sale_price and appraised_value is in thousands of dollars.

(b) Interpret the y-intercept of the least squares line. Does it have a practical meaning for this application? Explain.

Mathematically, the y-intercept means that when the property has no appraisal value, the sale price is predicted to be $1,358.68. However, in this case it has no practical application because the appraisal value cannot be zero and the predicted sale price of $1,358.69 also doesn't make much sense.

(c) Interpret the slope of the least squares line. Over what range of *x* is the interpretation meaningful?

When the appraised property value increases by $1000, the predicted sale price is expected to increase by $1,408.27.

In class Activity 3

The interpretation of the simple linear regression is only meaningful when the x value is inside the range of [113.1, 1262.3], outside which it becomes extrapolation.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   113.1   170.4   229.2   305.6   330.8  1262.3
```

(d) Use the least squares model to estimate the mean sale price of a property appraised at $300,000.

New appraised value = 300000 / 1000 # adjusting for units
New predicted sale price = 1.35868 + 1.40827 * 300 = 423.8397

The predicted mean sale price of property appraised at $300,000 is $423,839.70.

(e) Compute an estimate of σ.

```
## Analysis of Variance Table
##
## Response: yvariable
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## xvariable  1 6874034 6874034  1454.1 < 2.2e-16 ***
## Residuals 74  349833    4727
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here, SSE = 349833 and Df = 74
Estimate of σ = sqrt(SSE / Df) = 68.75662
Adjusting for the units, the estimated σ is $68,756.62.

Activity 3_2

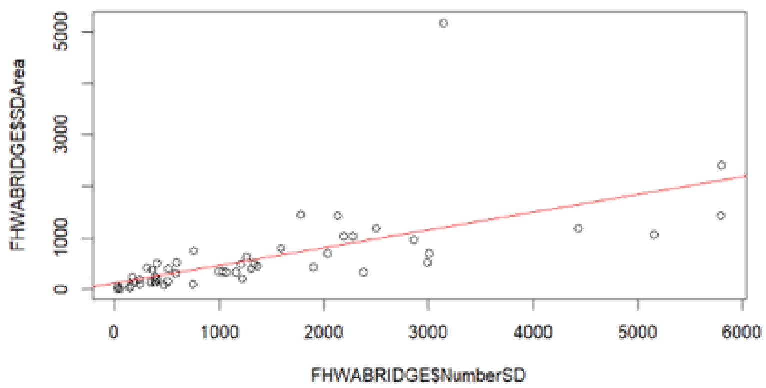
1. a.

```
#
# Coefficients:
#                 Estimate Std. Error t value Pr(>|t|)
# (Intercept)    119.86392  123.02005   0.974    0.335
# data1$NumberSD   0.34560    0.06158   5.613 8.69e-07 ***
#
```

Predicted_SDArea = 119.86392 + 0.34560 * NumberSD

b.

In class Activity 3



c. Check the 4 model assumptions on page 7 of the lecture chapter 3.

d.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 119.86392  123.02005   0.974    0.335
NumberSD      0.34560    0.06158   5.613 8.69e-07 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 635.2 on 50 degrees of freedom
```

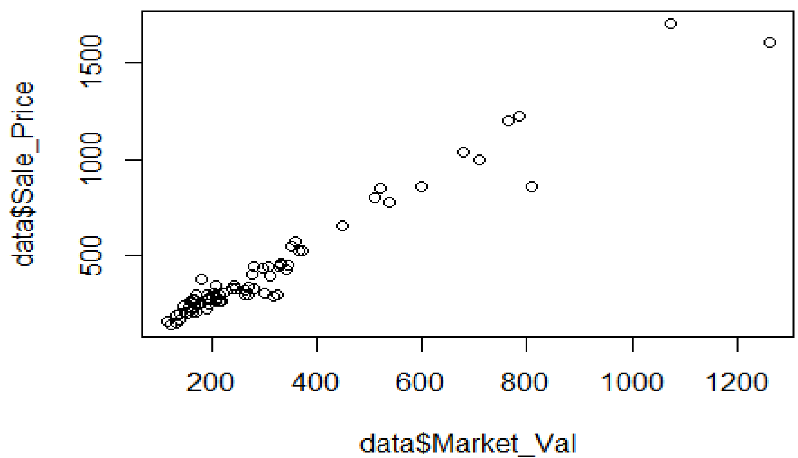
Or calculate it (Same as 1e)

```
## Residuals      50 20173064   403461
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

s <- sqrt(20173064/(52-2))
s

## [1] 635.186
```

e. $\hat{y} \pm (2 * 635.2)$ = $\hat{y} \pm 1270.4$

In class Activity 3

2a.



Since the scatterplot exhibits positive and linear relationship, we can run a least squares regression equation.

The results is as follows:
Predicted Sale_Price = 1.35868 + 1.40827 * Market_Value
(all the variables are in thousand of dollars)

For the hypothesis test:
H0: β1 = 0
Ha: β1 > 0

The p-value in the model output is for Ha: β1 ≠ 0.
P-value for Ha: β1 > 0 is < 0.01 (p-value/2), we conclude that at 0.01 level of significance we have enough evidence to reject the null hypothesis. We conclude that the slope (β1) of the straight line model is positive.

b.

```
                  2.5 %    97.5 %
(Intercept) -26.075006 28.792368
Market_Val    1.334683  1.481858
>
```

For every $1,000 increase in market value, we are 95% confident that the predicted average sale price will increase between $1,334.68 and $1,481.86
or

In class Activity 3

We are 95% confident that the sale price will increase between $1335 and $1482 for every $1000 increase in market value.

c. To obtain a narrower confidence interval we need to minimize the Standard Error by increasing the sample size.