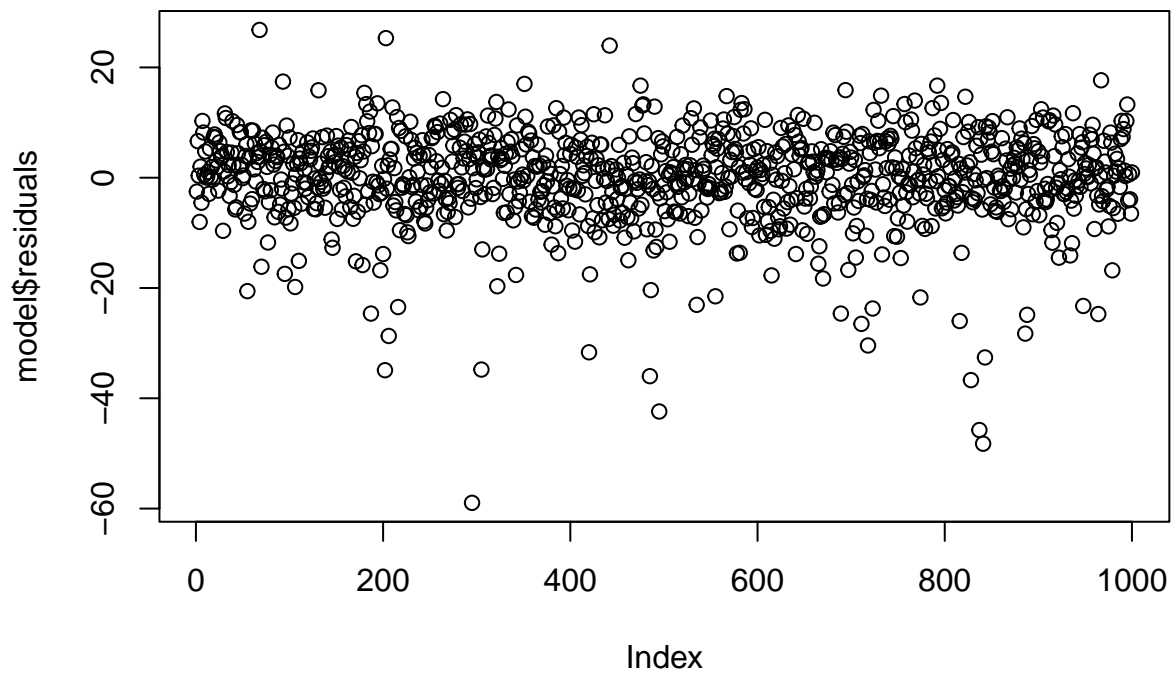


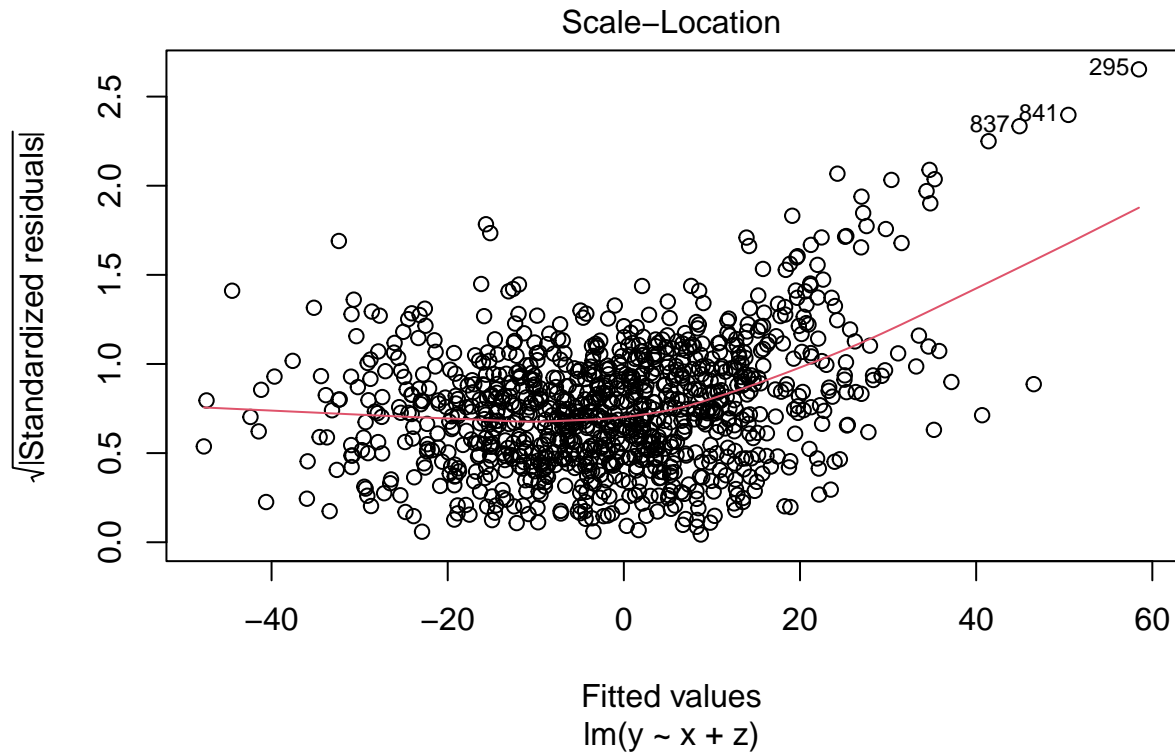
## hw 4

2023-11-15

### Problem 1

#1.1 #The assumption of homoskedasticity is that all the error terms or the disturbance in the relationship between the independent and dependent variable is the same in all the values of the independent variable. And the residuals all have the same variance for every value of fitted values and predictors. The consequences of violating homoskedasticity is when heteroscedasticity which is when the size of the error term is different across the values of an independent variable. This will skew the results of the test and make it biased. Also, if the errors are heteroskedastic, we won't be able to trust the standard errors of the estimates or be able to calculate it accurately.



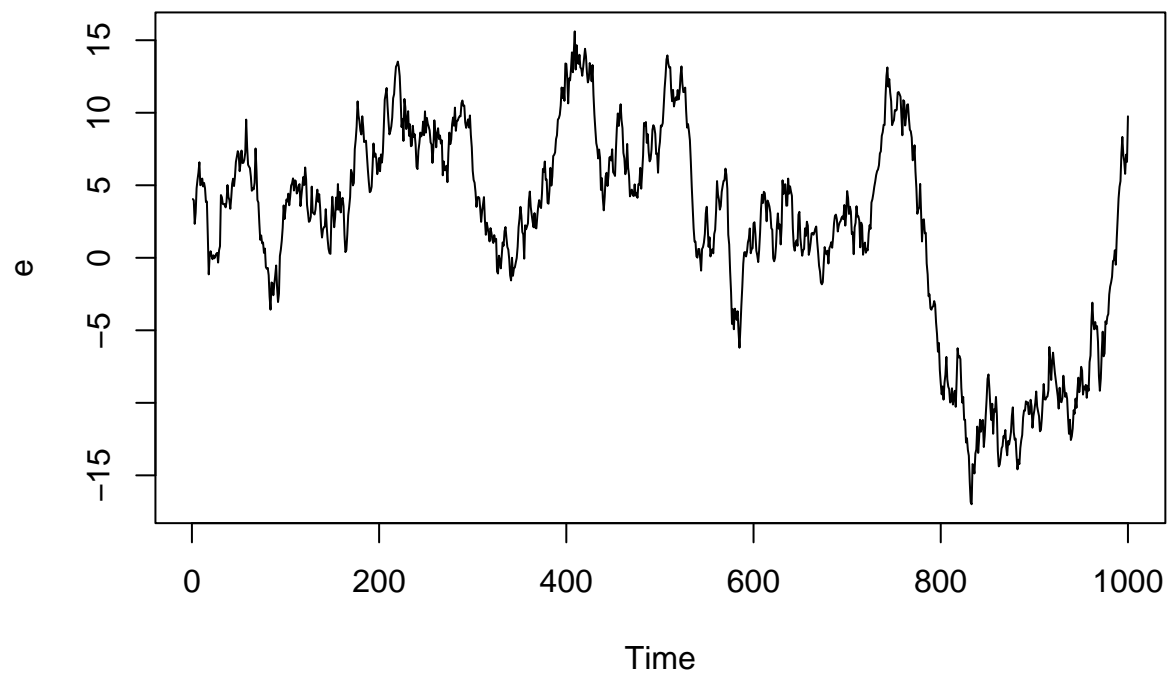


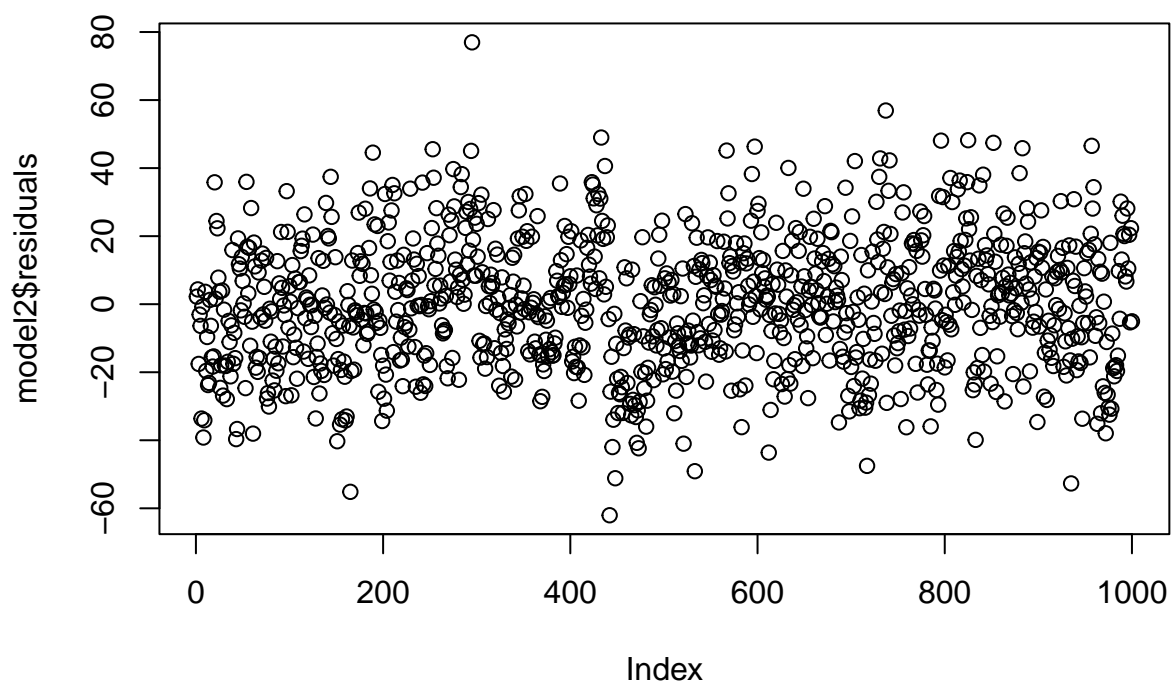
```
##
## studentized Breusch-Pagan test
##
## data: model
## BP = 158.27, df = 2, p-value < 2.2e-16
```

#1.2 #The residuals appear to be homoskedastic because the errors don't deviate too much from the flat line. The null is that there is homoskedasticity of errors and from the Breusch-Pagan test, we can see that with a p-value of 0.07038, we can't reject the null and therefore conclude that our residuals are homoskedastic; the assumption of homoskedasticity hasn't been violated.

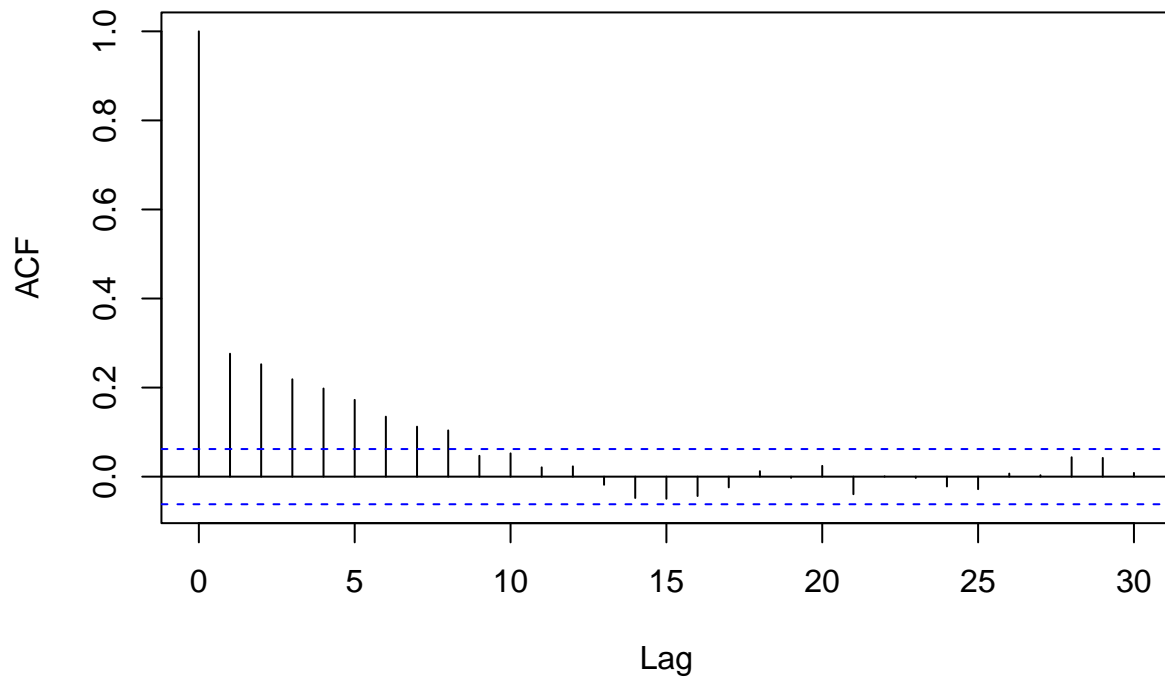
#1.3 #I don't have any concerns about the model I ran because the model appears to be homoskedastic and the assumption doesn't appear violated.

#Problem 2 #2.1 #The assumption of no auto correlation has to do with the fact that the error terms from dissimilar observations should not be correlated with each other. If violated, the OLS won't be blue and we won't see it as reliable and efficient (relating to the least squares estimates). Another violation will also produce wrong standard errors for the coefficients of the regression estimates.





## Series model2\$residuals



```
##
## Durbin-Watson test
##
## data: model2
## DW = 1.4478, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0

##
## Breusch-Godfrey test for serial correlation of order up to 3
##
## data: model2
## LM test = 123.31, df = 3, p-value < 2.2e-16
```

#The residuals appear to be autocorrelated because from plotting the residuals against time, we can't see the randomness and white noise. Our model demands randomness and white noise. From the acf model, we also see the random lags. From the Durbin-Watson test, we have a p-value less than pretty less than zero and the same for the Breusch Godfery test. We can reject the null of no serial correlation based on the tests we've ran. There is auto correlation because we can see a pattern from the lag graph. Also from the Durbin Watson test, we can see that the p-value is very small ( $2.2e-16$ ) which means we reject the null that there is no autocorrelation.

#2.3 I do have any concerns about the model I ran given the evaluation of the autocorrelation above because the assumption does appear violated. And we'd have to fix it.

[illegible]























































































[illegible]

















[illegible]

















```

## min; returning Inf

## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf

## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf

## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf

## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf

## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf

## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf

## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf

## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf

## [1] 0.002106718

## [1] 1.001495

## [1] -0.0004478986

## [1] 0.9762546

## [1] 0.004415461

## [1] 0.9501454

## [1] 0.003121808

## [1] 0.9385923

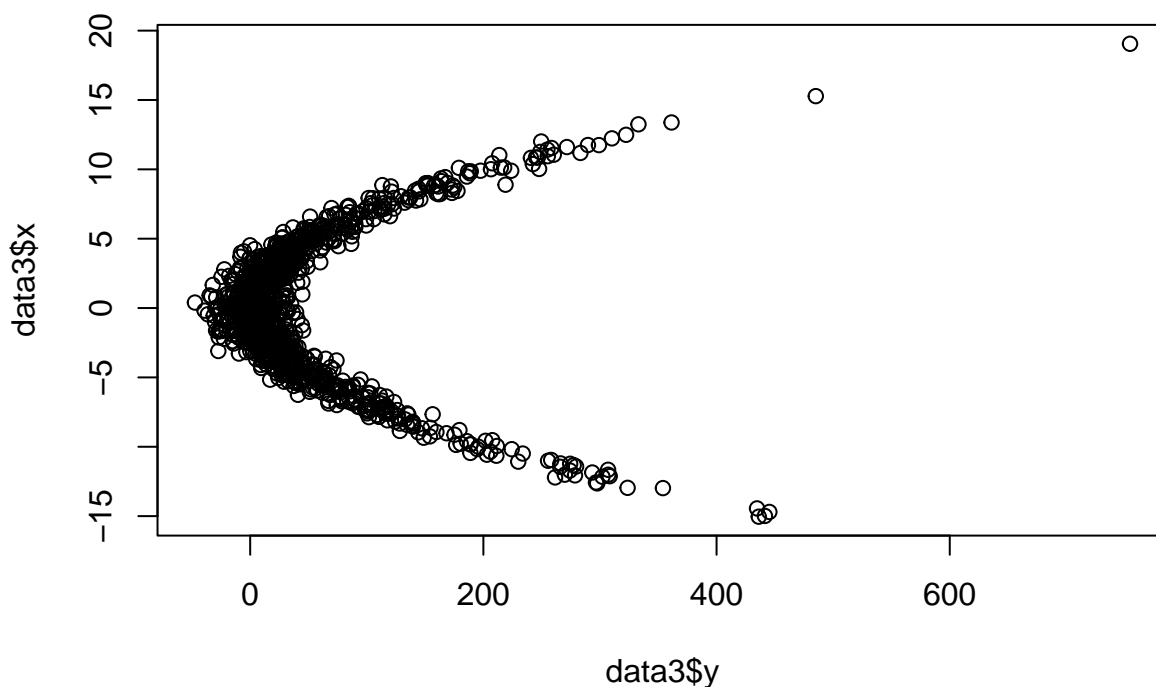
## [1] -0.009367948

## [1] 1.083753

```

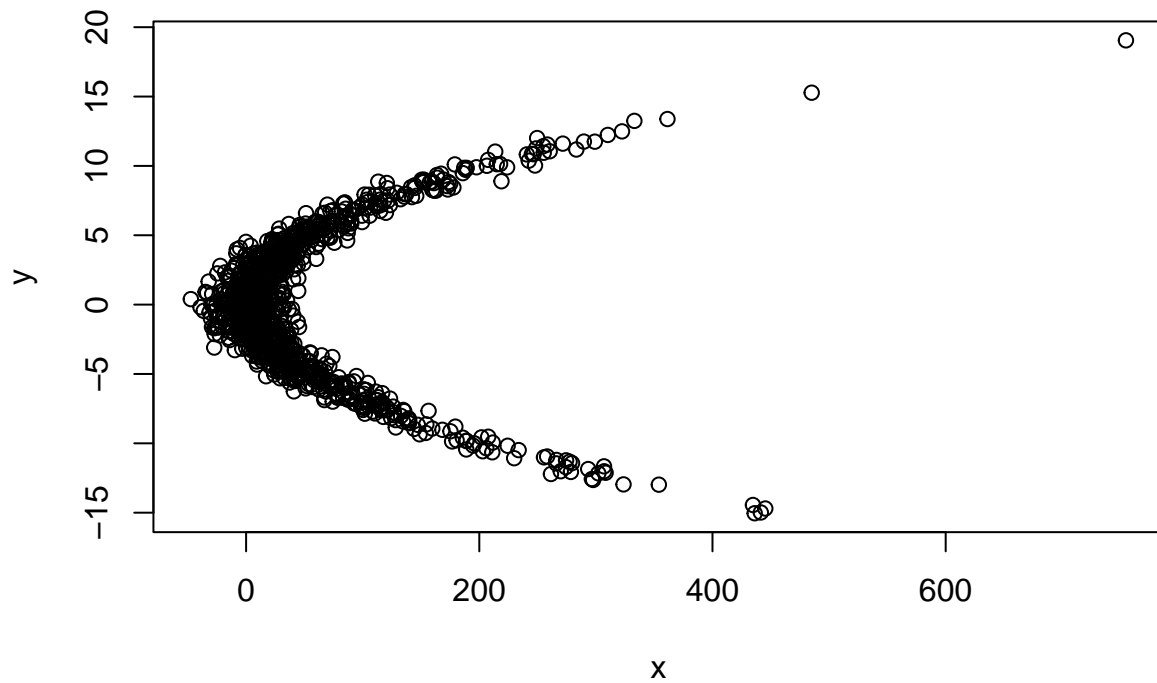
#1. The beta hat does vary across values of p but doesn't vary dramatically. We are aware that for the OLS to be blue, the beta shouldn't vary across (be unbiased) but here we do see a variance but not too much from each other. Even though they vary all of the sd's are closer to 1 and the mean is closer to zero. #2. Comparing my standard deviations of my estimates to their standard errors across different values of p, we do see that we have a small se which will affect the ratio. Therefore, the standard errors will overestimate the true variability of the estimates.

#in the presence of heteroskedasticity, the least squares estimator b is still unbiased, consistent and asymptotically normally distributed



#Problem 3

## Model Showing Relationship Between x and y



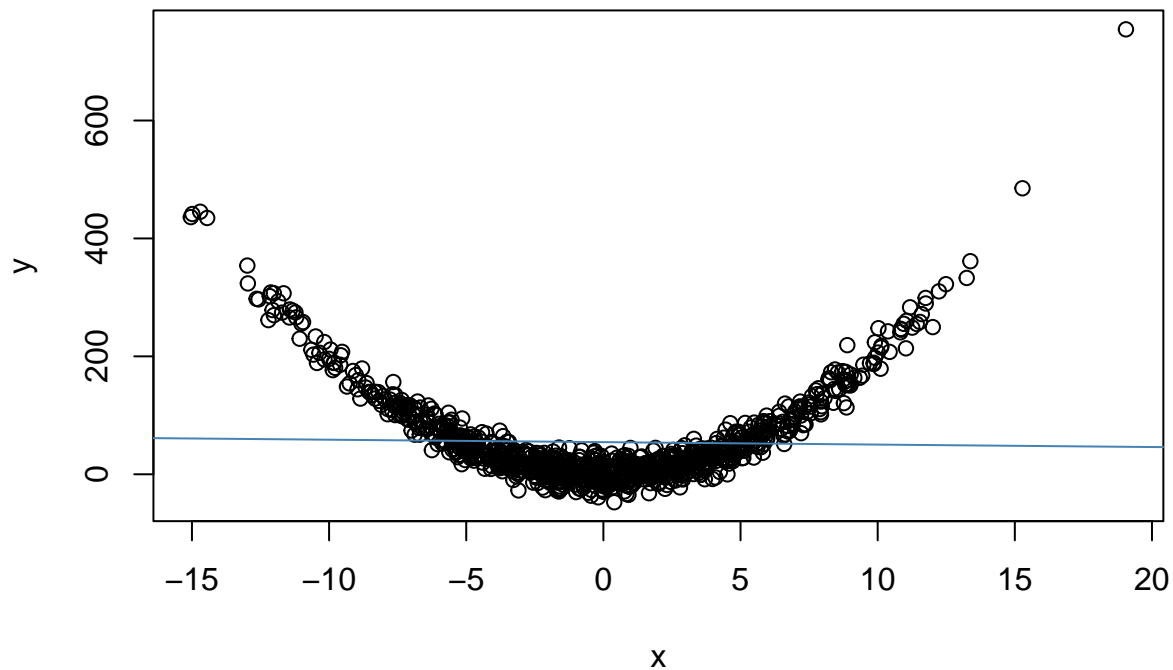
```
## integer(0)
```

```
#3.1 # The model doesn't appear to fit the observed relationship at all because it doesn't look like it's a  
perfect line of best fit for the model.
```

```
#3.1
```

```
## integer(0)
```

```
## Warning in abline(reg3, col = "steelblue"): only using the first two of 3  
## regression coefficients
```

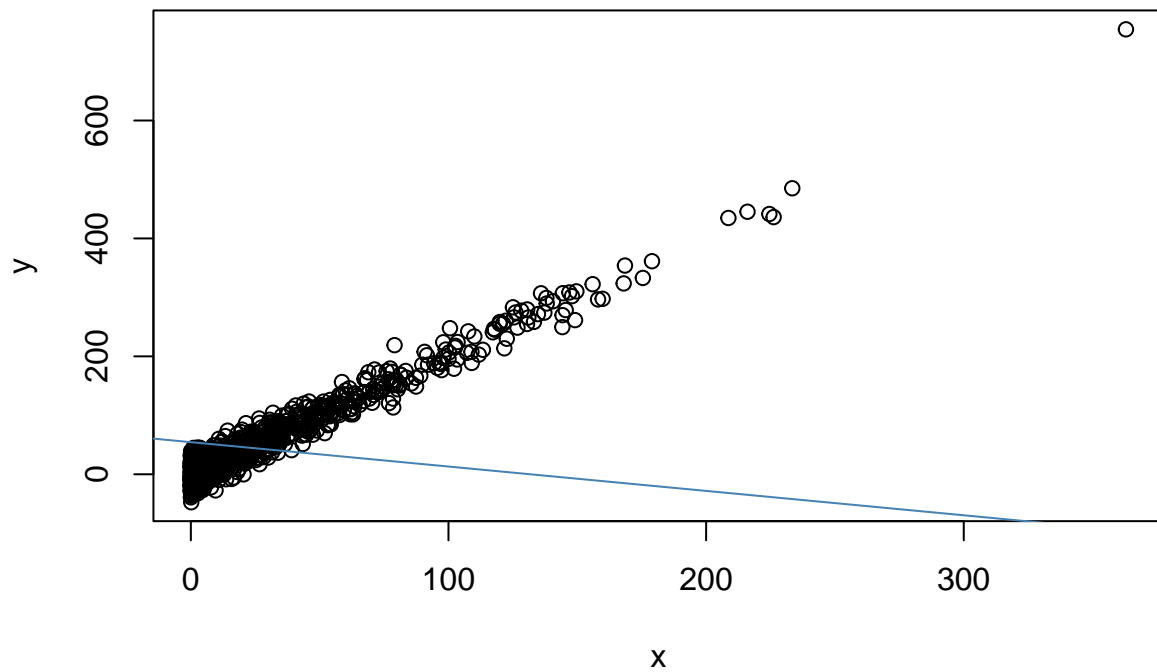


```
##
## Call:
## lm(formula = y ~ x + z, data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -66.99 -46.54 -29.89  13.00 685.04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   54.5003     2.4072  22.641  < 2e-16 ***
## x             -0.4135     0.4653  -0.889   0.374
## z              3.4251     0.4631   7.396 2.97e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 76.11 on 997 degrees of freedom
## Multiple R-squared:  0.05265,    Adjusted R-squared:  0.05075
## F-statistic: 27.71 on 2 and 997 DF,  p-value: 1.948e-12

## integer(0)

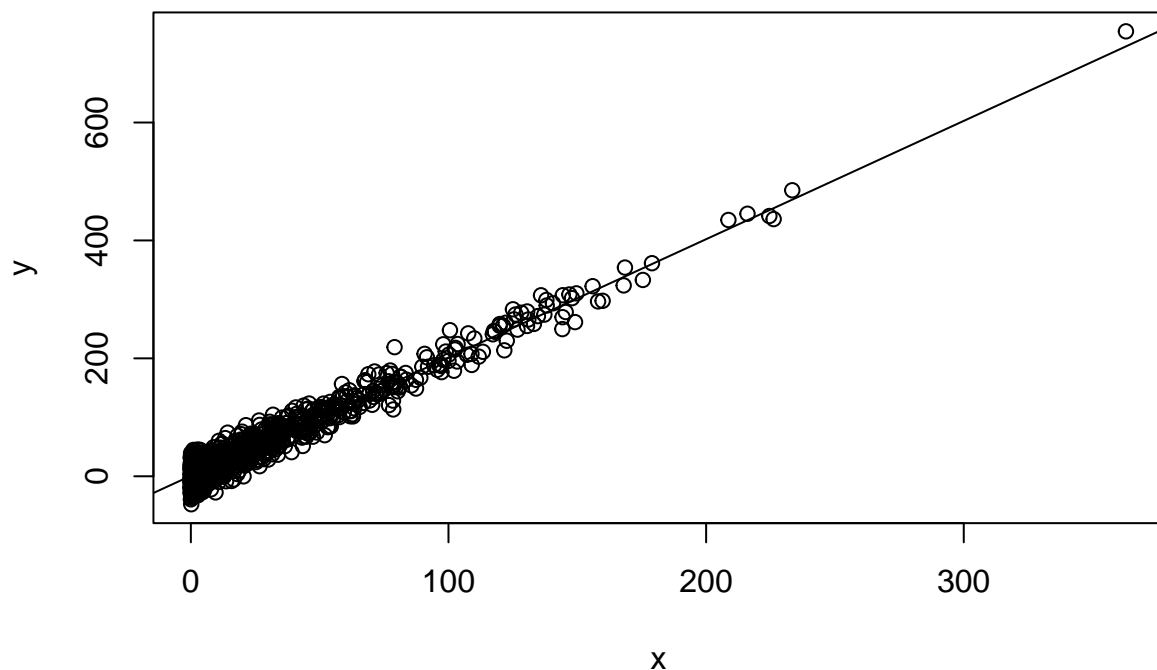
## Warning in abline(reg3, col = "steelblue"): only using the first two of 3
## regression coefficients
```





#From the r square, we can see that the r square is less than 5 percent so it doesn't fit the model at all.

## Warning in abline(reg4): only using the first two of 3 regression coefficients



#Now the model appears to fit the observed relationship better after squaring the x and we do reach a different conclusion with this new model and fit (best fit)

```
##
## Please cite as:

## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

##
## All models
## =====
##                               Dependent variable:
##                               -----
##                               y
##                               (1)      (2)
## -----
## x                               -0.413
##                               (0.465)
##
## x.squared                       2.007***
##                               (0.004)
##
## z                               3.425***      3.020***
##                               (0.463)      (0.029)
```

```
##
## Constant          54.500***    0.811***
##                  (2.407)      (0.187)
##
## -----
## Observations      1,000        1,000
## R2                0.053        0.996
## Adjusted R2       0.051        0.996
## Residual Std. Error (df = 997) 76.107    4.838
## F Statistic (df = 2; 997)    27.707*** 129,728.500***
## =====
## Note:              *p<0.1; **p<0.05; ***p<0.01
```