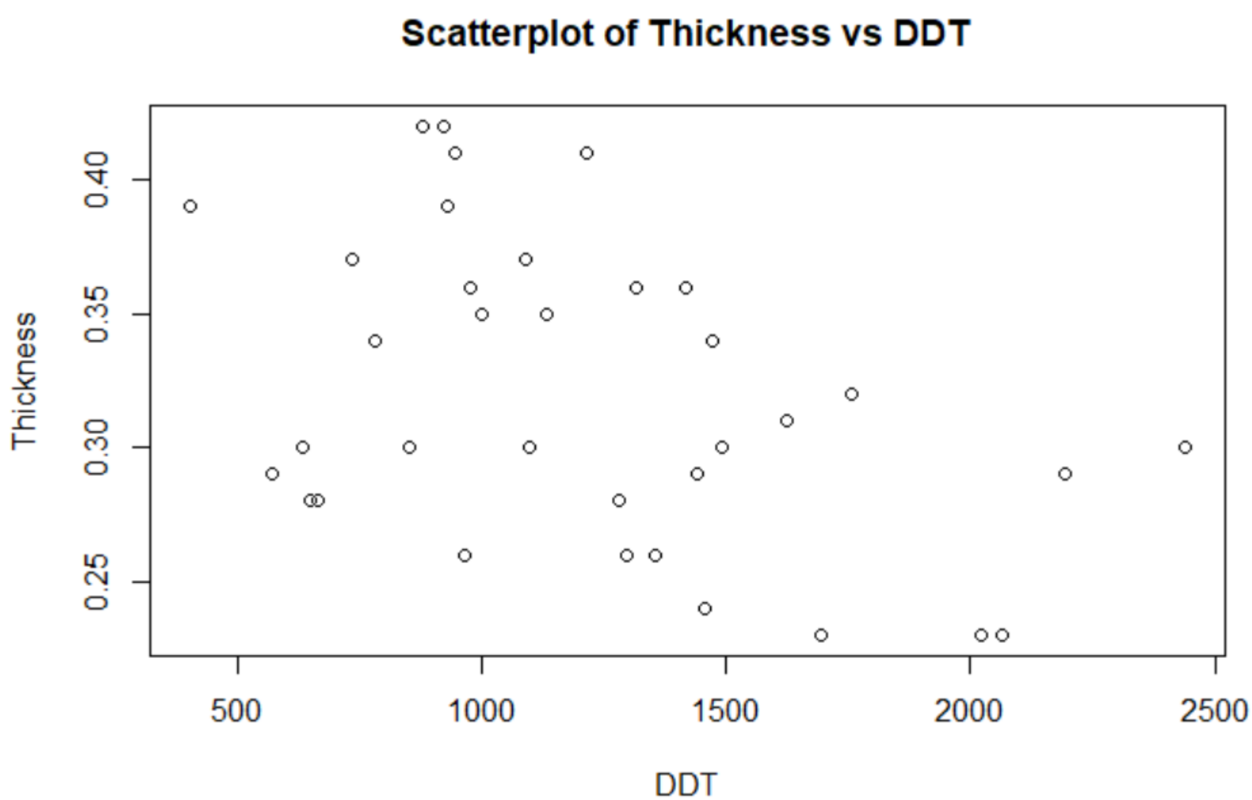
Stat*3240, Assignment 1
Graham Eckel
0679576

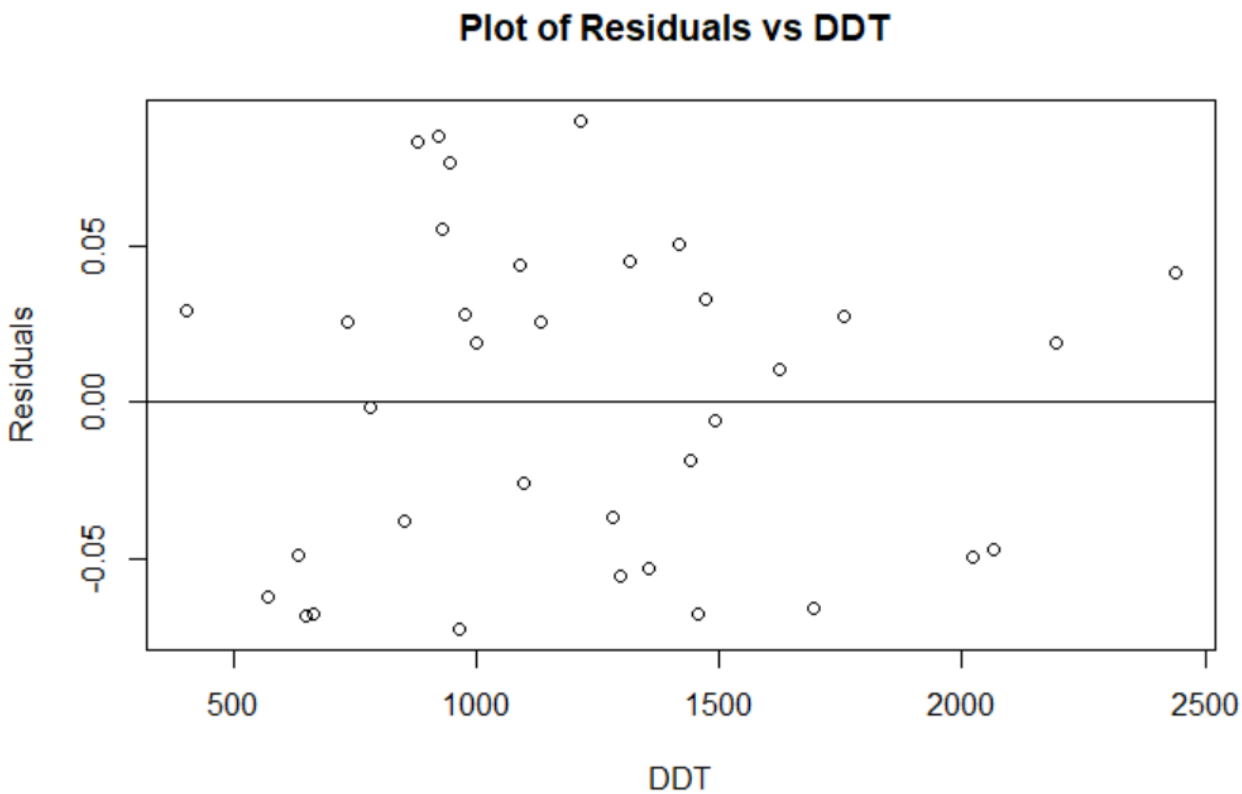
Q1
a)
I've chosen to do this question in R instead of SAS

```
set.seed(2019-09-12)
dir = "C:\\Users\\...\\Applied Regression Analysis\\"
file1 = "3240_F19_A1_DDT.txt"
dfDDT = read.table(file=paste(dir,file1, sep=""), header=TRUE, sep=' ')

plot(dfDDT$DDT, dfDDT$thickness, ylab = "Thickness", xlab = "DDT", main = "Scatterplot of
Thickness vs DDT")
```



```
resThick = resid(lmThick)
plot(dfDDT$DDT, resThick, ylab = "Residuals", xlab = "DDT", main = "Plot of Residuals vs DDT
")
abline(0,0)
```



lmThick = lm(dfDDT$thickness~dfDDT$DDT)
summary(lmThick)

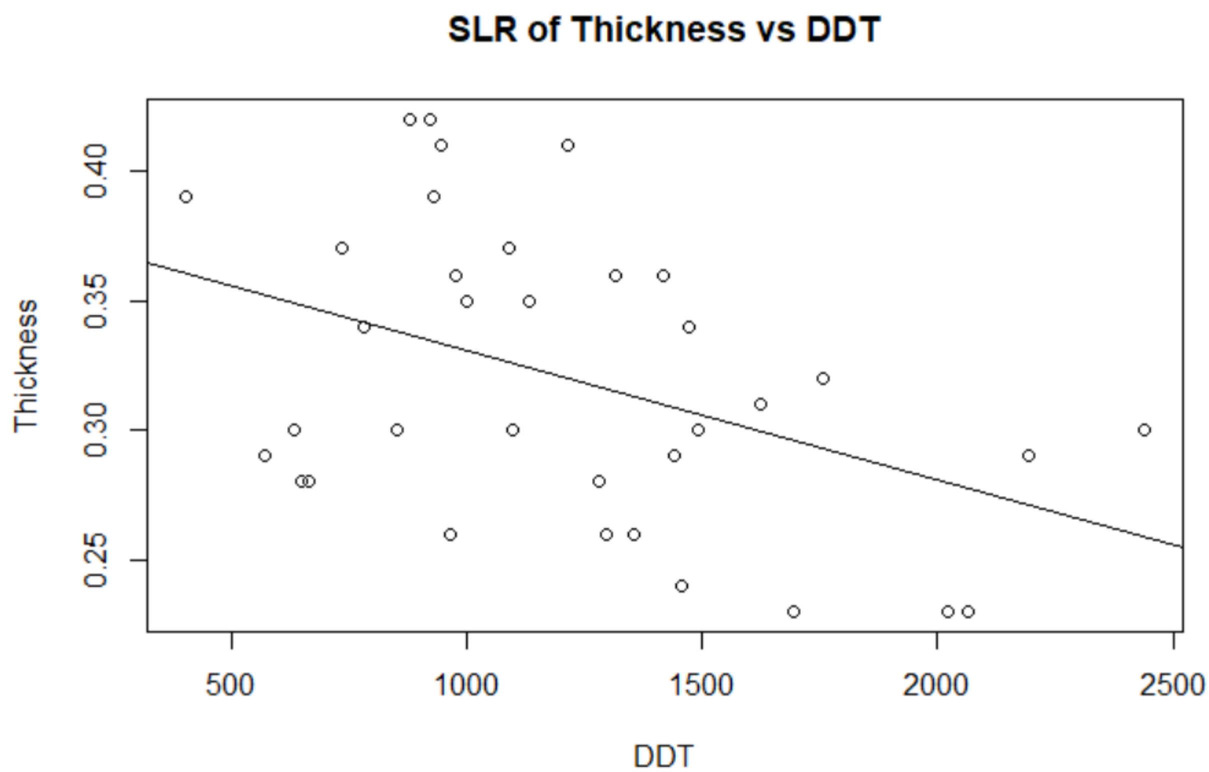
```
Call:
lm(formula = dfDDT$thickness ~ dfDDT$DDT)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.07255 -0.04945  0.01040  0.03718  0.08989 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.810e-01  2.412e-02  15.791   <2e-16 ***
dfDDT$DDT   -5.016e-05  1.840e-05  -2.726   0.0102 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05204 on 33 degrees of freedom
Multiple R-squared:  0.1838,	Adjusted R-squared:  0.1591 
F-statistic: 7.432 on 1 and 33 DF,  p-value: 0.01018
```

plot(dfDDT$DDT, dfDDT$thickness, ylab = "Thickness", xlab = "DDT", main = "SLR of Thickness vs DDT")
abline(lmThick)



b)
In this context, the parameter estimate of the y intercept can be interpreted to mean the thickness of the eggshells when there is no DDT present. The slope can be interpreted to mean the average unit of thickness decrease for each unit increase of DDT.

c)
Threshold t-value = qt(0.975, df = 33) = 2.0345

We are testing to see if there exists a linear relationship between thickness and DDT. If there is no relationship, we will have a flat regression line with a slope of zero. So, we set this as our null hypothesis.

Ho: B1 = 0	- null hypothesis:	predicted slope, B1 is equal to 0
Ha: B1 =/= 0	- alternative hypothesis:	predicted slope, B1, is not equal to 0

At a 5% significance, our threshold t-value for a two tailed test with 33 degrees of freedom is 2.0345. Since the absolute value of our t-value from our observations is 2.726 - which is greater than our threshold value - we have evidence to reject the null hypothesis. It can also be said that since our p value of 0.0102 is less than our significance of 0.05, we have evidence to reject the null. In conclusion, we reject the null hypothesis that there is no linear relationship between eggshell thickness and DDT contaminants since it can be seen that there is strong evidence that as the amount of DDT increases, the thickness of the eggshells decreases.

d)
A 95% confidence interval for the true slope can be calculated as such:

Threshold t-value	= qt(0.975, df = 33) = 2.0345
Standard Error of of Slope	= 0.0000184
B1hat	= -0.00005016

B1hat +- Standard Error of of Slope*Threshold t-value =(-8.759032e-05, -1.2725e-05)

Or, simply using the r function:

confint(lmThick, "dfDDT$DDT", level=0.95) = (-8.759032e-05, -1.2725e-05)

e)
Threshold t-value = qt(0.975, df = 33) = 2.0345

We are testing to see if the true y intercept is zero. So, we set this as our null hypothesis.

Ho: B0 = 0	- null hypothesis:	predicted y intercept, B0, is equal to 0
Ha: B0 =/= 0	- alternative hypothesis:	predicted y intercept, B0, is not equal to 0

At a 5% significance, our threshold t-value for a two tailed test with 33 degrees of freedom is 2.0345. Since our t-value from our observations is 15.971 - which is greater than our threshold value - we have evidence to reject the null hypothesis. It can also be said that since our p value of 2e-16 is less than our significance of 0.05, we have evidence to reject the null. In conclusion, we reject the null hypothesis that the y intercept is zero. There is strong evidence that our SLR does not have a y intercept of zero.

This is not a particularly meaning test.

f)
The distribution of points around the midline at (0,0) does not seem to be decreasing or increasing as DDT increases. This indicates the assumption of common variance is holding. As well, the assumption of homoscedasticity appears to hold as the data does not appear to be ‘curving’ around the midline.