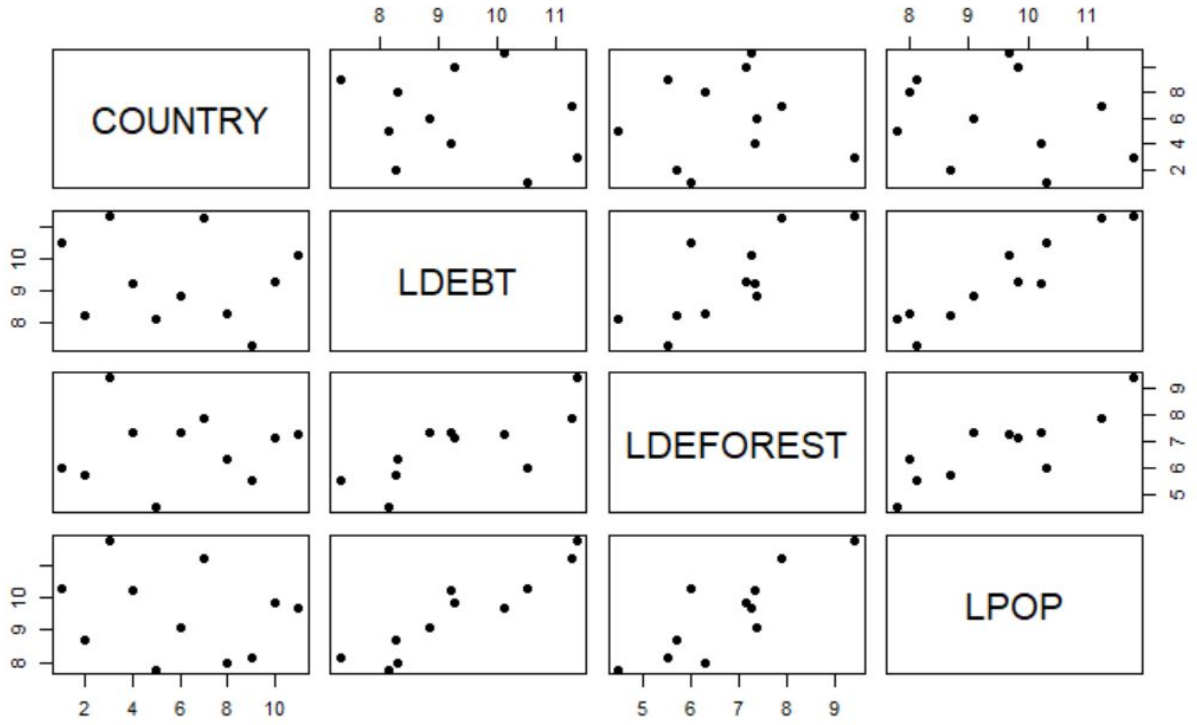
Assignment 4
Part 2

```
dir = "..."
file1 = "deforest.csv"
dfDeforest = read.table(file=paste(dir,file1, sep=""), header=TRUE, sep=',')
```

2a)
We are tasked with creating a scatterplot matrix for the deforest dataset. Then, comment on the *appearance* of linearity of three of the four dimensions: LDEBT, LDEFOREST, LPOP. We are *not* asked to reduce the scatterplot matrix to three dimensions, test for covariance/correlation, or to test hypothesis.

```
pairs(dfDeforest, pch = 19)
#pairs(dfDeforest[,2:4], pch = 19)
```



For the entire matrix, If the x-axis is A B C D (left to right) and the y axis is 1 2 3 4 (bottom to top), then LDEBT by LDEFOREST is (3,C) where LDEBT is on the y-axis and LDFOREST is on the x-axis. This relationship appears to be positive and relatively linear. LDEBT by LPOP (3,D) is positive and relatively more linear than LDEBT by LDEFOREST. Lastly, LDEFOREST by LPOP (2,D) appears to positive and relatively less linear than LDEBT by LPOP, but still linear. The remaining three lower panel boxes are simply mirror images, flipped axis, of the three already discussed.

2b)

```
LDEBT = dfDeforest[,2]
LDEFOREST = dfDeforest[,3]
LPOP = dfDeforest[,4]
M1 = lm(LDEFOREST~LDEBT)
summary(M1)
```

Table1

```
Call:
lm(formula = LDEFOREST ~ LDEBT)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6595 -0.2926  0.2844  0.5185  1.1096

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.2512     2.0479  -0.123   0.9051
LDEBT         0.7519     0.2175   3.457   0.0072 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

It is implied, but not explicitly stated, that we should test a hypothesis for LDEBT on LDEFOREST. This exact scenario happened in question 1c) from assignment 3, in which I was marked incorrect for not including an hypothesis test. So, tests incoming!

5 Step Hypothesis test:

1. H0: b1 = 0, The mean of LDEFOREST is not a function (independent) of LDEBT
   H1: b1 =/= 0, The mean of LDEFOREST is a function of LDEBT, SLR is adequate

2. Significance level, a = 0.05

3. From Table 1 we can see that our p value for the slope, B1 = 0.0072

4. We will reject the null hypothesis if p value < significance. We can see that 0.0072 is < 0.05 so we have evidence to reject the null.

5. There is evidence to reject the null hypothesis, that the reduced model is the best fit for this data.

2c)
It is implied, but not explicitly stated, that we should test a hypothesis:

Table2

```
Call:
lm(formula = LDEFOREST ~ LDEBT + LPOP)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3342 -0.3330 -0.2078  0.5914  0.9583

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.3856     1.8527  -0.748   0.4760
LDEBT        -0.2208     0.5156  -0.428   0.6797
LPOP          1.0722     0.5294   2.025   0.0774 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7967 on 8 degrees of freedom
Multiple R-squared:  0.716,     Adjusted R-squared:  0.645
F-statistic: 10.09 on 2 and 8 DF,  p-value: 0.006504
```

5 Step Hypothesis test:

1. H0: b1 = 0, The reduced model is adequate
   H1: b1 =/= 0, After taking into account LPOP, the full model is necessary

2. Significance level, a = 0.05

3. From Table 1 we can see that our p value for the slope, B1 = 0.6797, after we include the effects of LPOP on LDEFORESTATION.

4. We will reject the null hypothesis if p value < significance. We can see that 06797 is < 0.05 so we do not have evidence to reject the null.

5. There is not evidence to reject the null hypothesis that LDEBT does not have a signficant effect on LDEFORESTATION after we include the affects of LPOP.

2d)
In our first model, the SLR, we found that as the debt of the nation increases, so too does their deforestation rate. However, once we included the growth of the nations population we can see that as the nation's population increased, so did the rate of deforestation, but, their debt and deforestation become inversely correlated. Since, in this case, LPOP is influencing both our dependent and independent variables, it is a confounding variable. Given this knowledge, a plausible explanation for why LDEBT switches from positive and significant to negative and not significant is that as the population increases, the rate of deforestation increases while the rate of debt decreases. Thus, if we include LPOP in our model, the correlation between LDEFORESTATION and LDEBT becomes inverted.