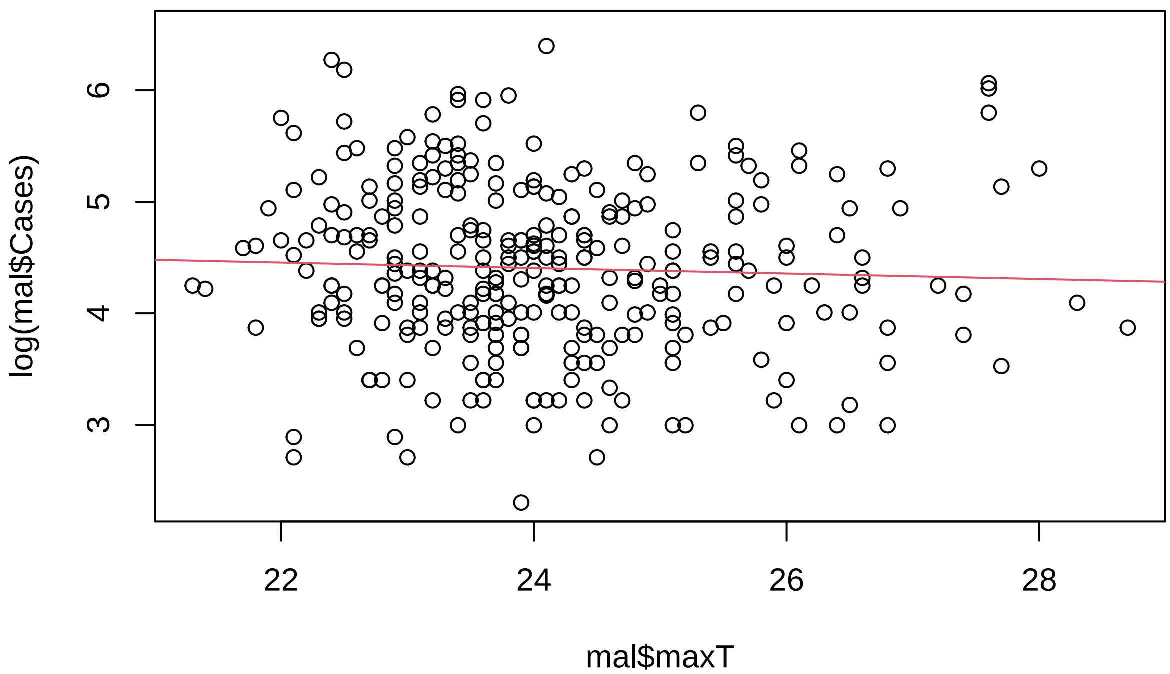
Time seried in epidemiology: Lab Exercises

After loading the `Kericho.csv` data in R. Please provide an answer to the following questions.

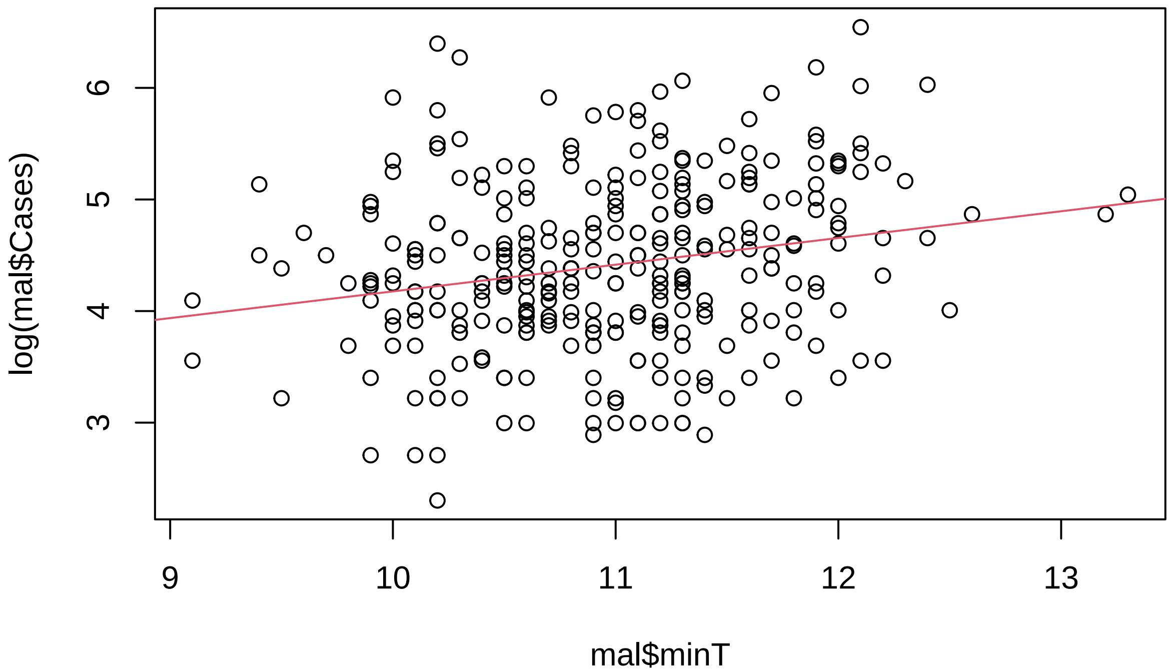
Exercise 1

1. Explore the association between maximum temperature (`maxT`), minimum temperature (`minT`) and rainfall (`Rain`) with the log-transformed number of reported malaria cases (`Cases`). What relationships do your observe and how strong are these?

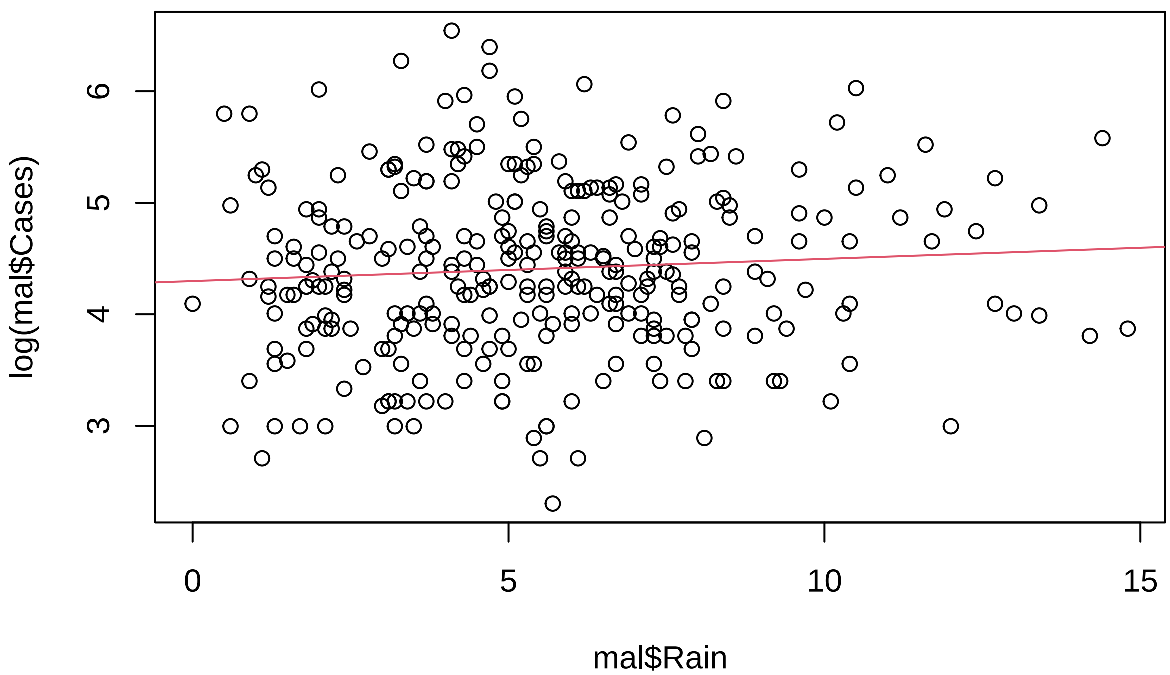
```
mal <- read.csv("Kericho.csv")

plot(mal$maxT,log(mal$Cases))
abline(lm(log(Cases)~maxT,data=mal),col=2)
```




```
plot(mal$minT,log(mal$Cases))
abline(lm(log(Cases)~minT,data=mal),col=2)
```




```
plot(mal$Rain,log(mal$Cases))
abline(lm(log(Cases)~Rain,data=mal),col=2)
```



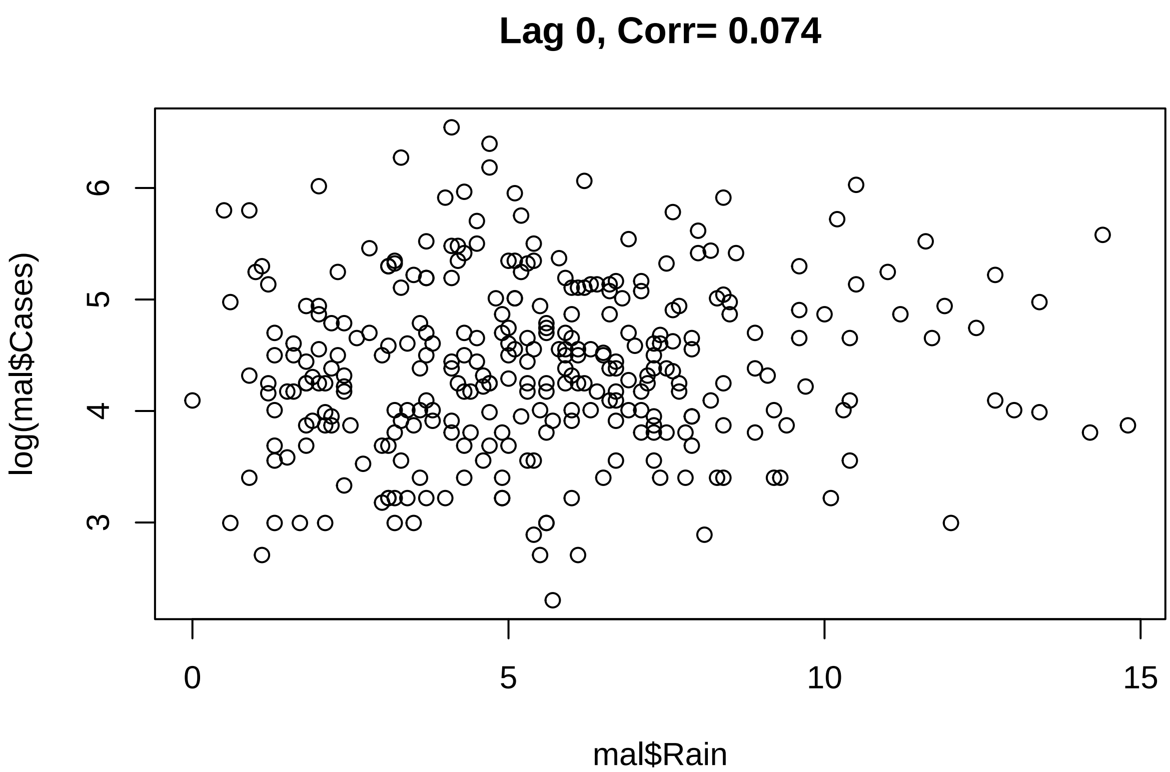
2. Consider the rainfall variable `Rain`. Create 4 time lagged variables that for a given month, give the rainfall amount of 1 month, 2 months, 3 months ago and 4 months ago. Plot the log-transformed number of cases against each of these variables. What do you observe?

```
mal$Month <- factor(mal$Month,levels =
                 c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep",
                   "Oct","Nov","Dec"),ordered=TRUE)
mal$t <- as.numeric(mal$Month)+12*(mal$Year-1979)
```

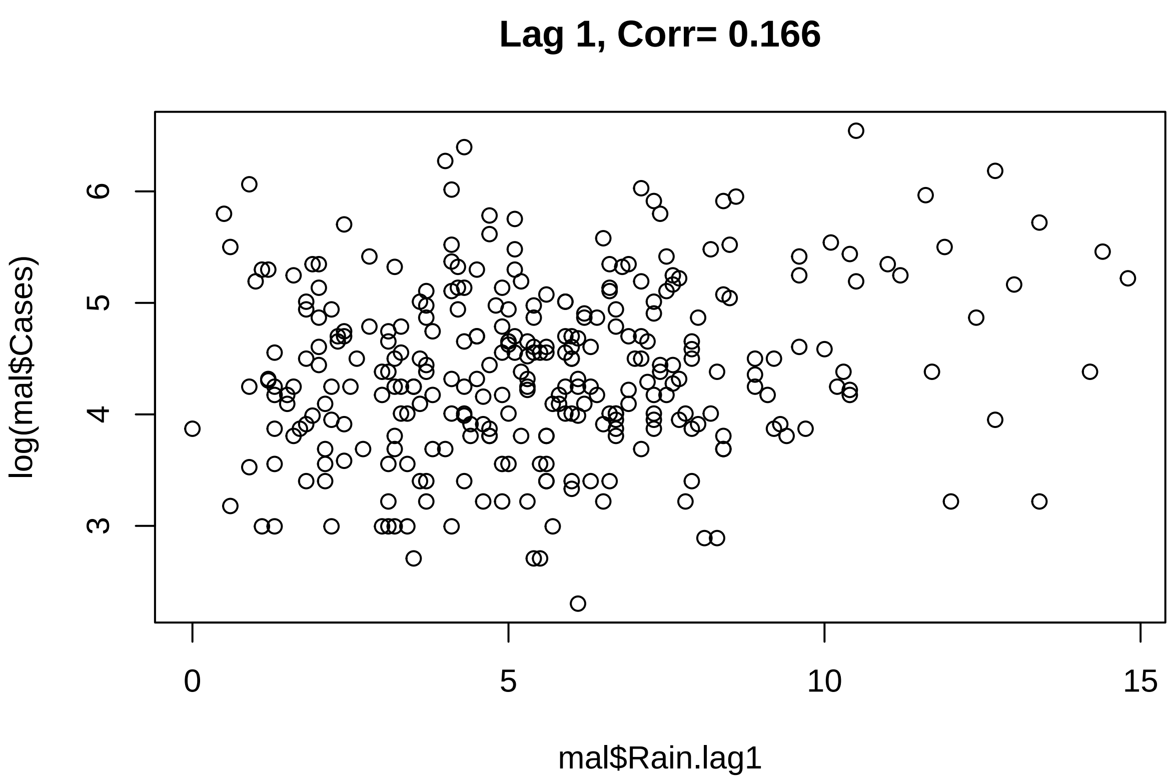
```r
create.lag.var <- function(x,lag) {
  tlag <- mal$t-lag
  ind.lag <- sapply(tlag,function(x) {
              out <- which((mal$t-x)==0)
              if(length(out)==0) out <- NA
              return(out)
             })
  x[ind.lag]
}


mal$minT.lag1 <-create.lag.var(mal$minT,lag=1)

# Check on newly created variable
head(mal[sort.list(mal$t),c("minT.lag1","minT","t")])
```

```
##   minT.lag1 minT t
## 1        NA 11.8 1
## 2      11.8 11.3 2
## 3      11.3 10.9 3
## 4      10.9 12.0 4
## 5      12.0 10.9 5
## 6      10.9 11.4 6
```

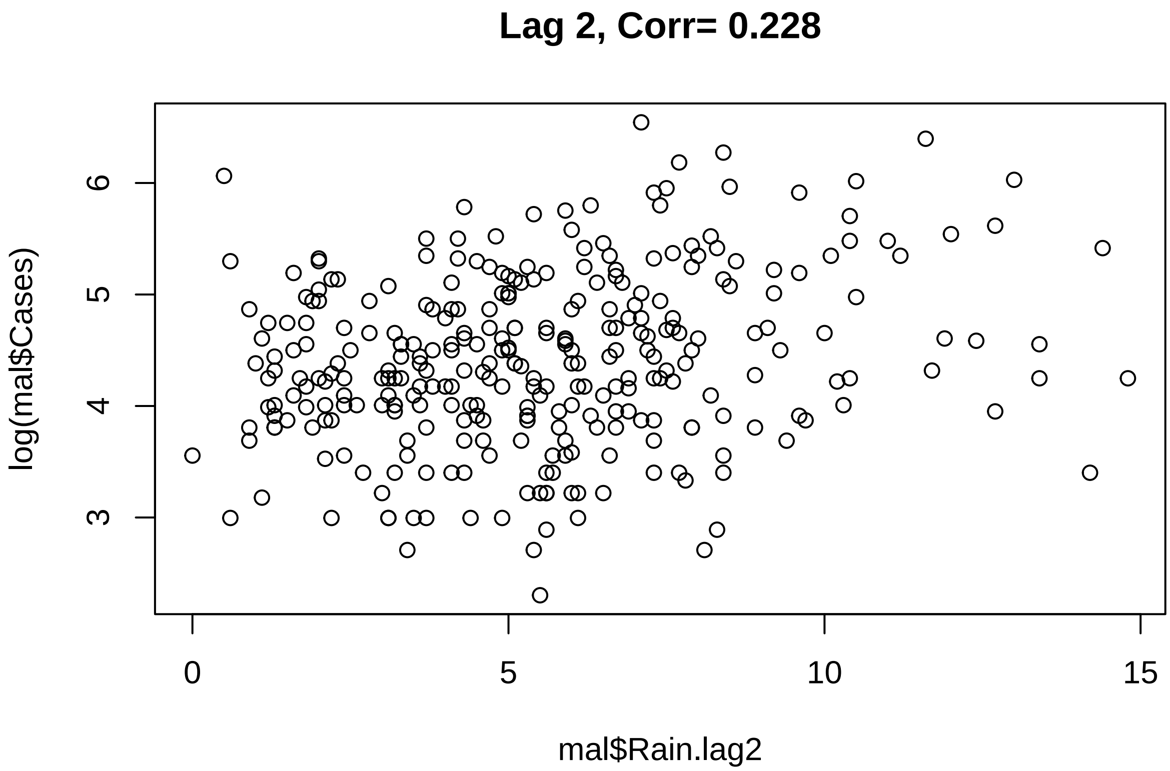
```r
mal$Rain.lag1 <- create.lag.var(mal$Rain,lag=1)
mal$Rain.lag2 <- create.lag.var(mal$Rain,lag=2)
mal$Rain.lag3 <- create.lag.var(mal$Rain,lag=3)
mal$Rain.lag4 <- create.lag.var(mal$Rain,lag=4)

plot(mal$Rain,log(mal$Cases),
     main=paste("Lag 0, Corr=",round(cor(mal$Rain,log(mal$Cases)),3)))
```



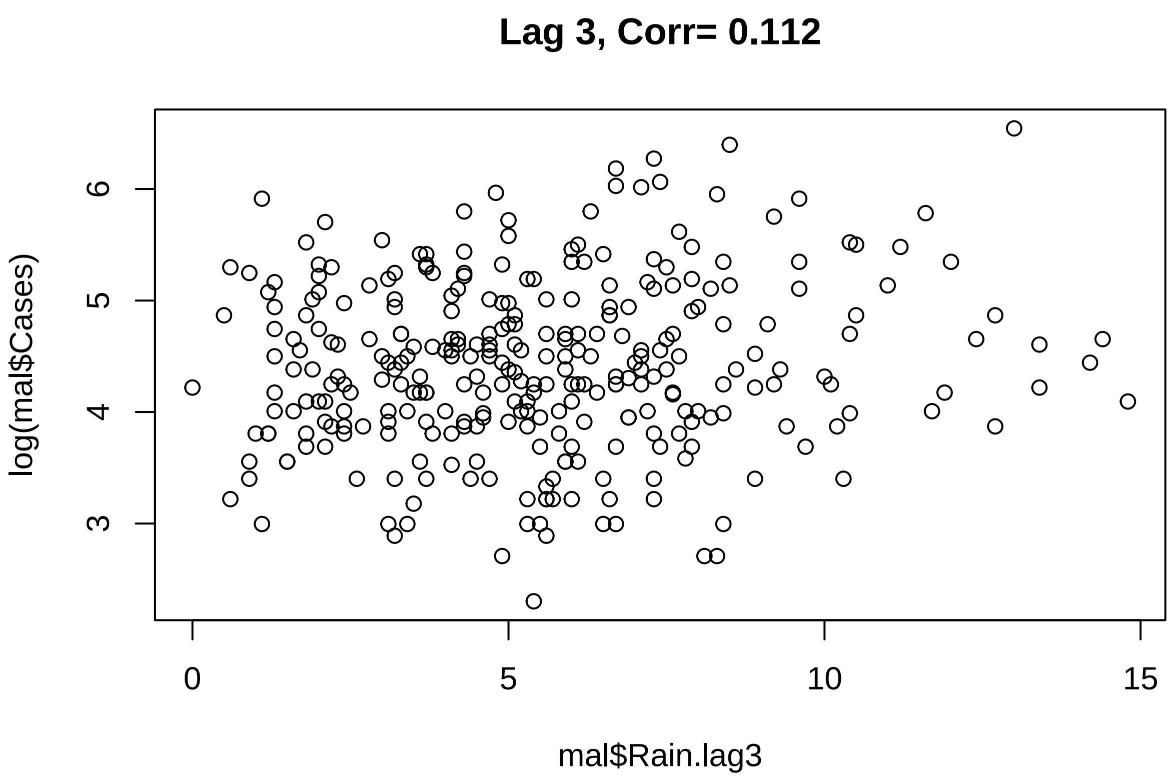

```r
plot(mal$Rain.lag1,log(mal$Cases),
     main=paste("Lag 1, Corr=",round(cor(mal$Rain.lag1,log(mal$Cases),use="complete.obs"),3)))
```



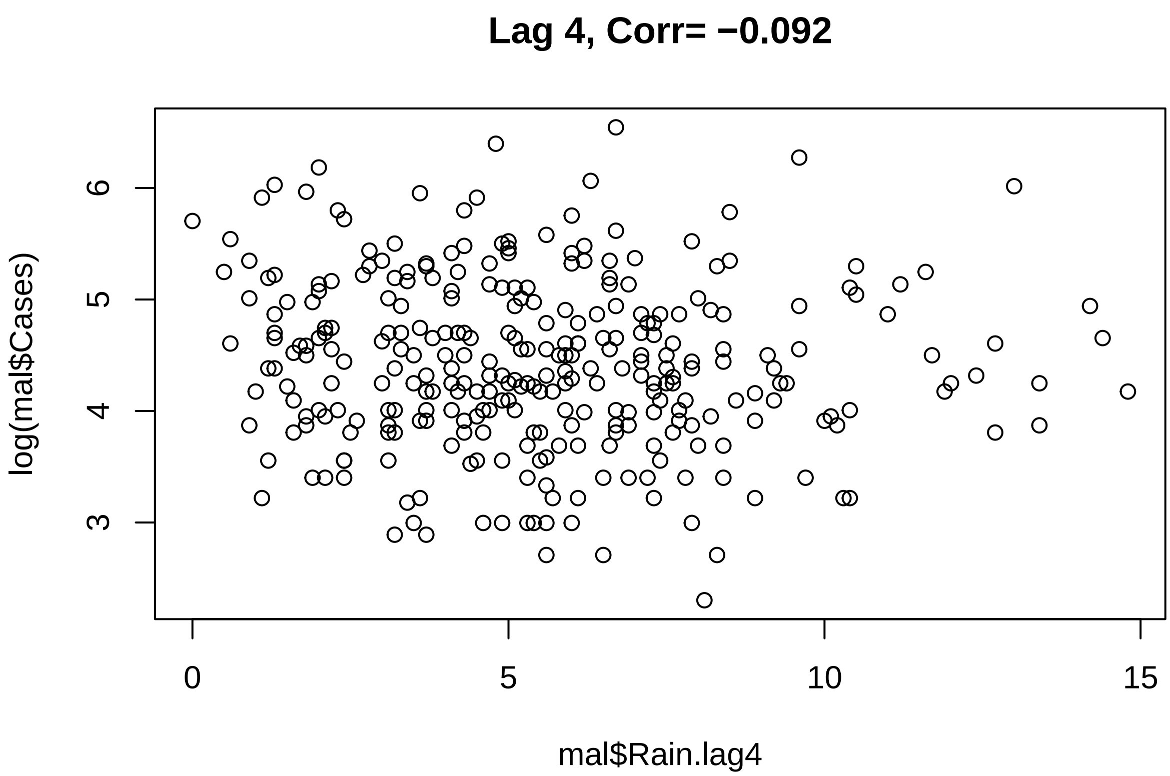

```r
plot(mal$Rain.lag2,log(mal$Cases),
     main=paste("Lag 2, Corr=",round(cor(mal$Rain.lag2,log(mal$Cases),use="complete.obs"),3)))
```




```
plot(mal$Rain.lag3,log(mal$Cases),
     main=paste("Lag 3, Corr=",round(cor(mal$Rain.lag3,log(mal$Cases),use="complete.obs"),3)))
```




```
plot(mal$Rain.lag4,log(mal$Cases),
     main=paste("Lag 4, Corr=",round(cor(mal$Rain.lag4,log(mal$Cases),use="complete.obs"),3)))
```



3. Fit a linear model for the log-transformed number of cases for each of the 4 lagged rainfall variables created in the previous point. Based on the fitted models, which lag has a stronger relationship with the reported malaria cases?

```
lm.lag0 <- lm(log(Cases) ~ Rain, data = mal)
lm.lag1 <- lm(log(Cases) ~ Rain.lag1, data = mal)
lm.lag2 <- lm(log(Cases) ~ Rain.lag2, data = mal)
lm.lag3 <- lm(log(Cases) ~ Rain.lag3, data = mal)
lm.lag4 <- lm(log(Cases) ~ Rain.lag4, data = mal)

summary(lm.lag0)
```

```
##
## Call:
## lm(formula = log(Cases) ~ Rain, data = mal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.10892 -0.49873 -0.03262  0.52464  2.16426
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.29802    0.09424  45.607   <2e-16 ***
## Rain         0.01991    0.01522   1.308    0.192
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7678 on 308 degrees of freedom
## Multiple R-squared:  0.005526,   Adjusted R-squared:  0.002297
## F-statistic: 1.712 on 1 and 308 DF,  p-value: 0.1918
```

```r
summary(lm.lag1)
```

```
##
## Call:
## lm(formula = log(Cases) ~ Rain.lag1, data = mal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.13518 -0.50605 -0.02626  0.53511  2.03922
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.16645    0.09336  44.626  < 2e-16 ***
## Rain.lag1    0.04448    0.01505   2.955  0.00337 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7575 on 307 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.02765,    Adjusted R-squared:  0.02448
## F-statistic:  8.73 on 1 and 307 DF,  p-value: 0.003372
```

```r
summary(lm.lag2)
```

```
##
## Call:
## lm(formula = log(Cases) ~ Rain.lag2, data = mal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.11249 -0.48276 -0.00874  0.55499  2.03159
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.08080    0.09198  44.367  < 2e-16 ***
## Rain.lag2    0.06078    0.01483   4.098 5.34e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7462 on 306 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.05202,    Adjusted R-squared:  0.04893
## F-statistic: 16.79 on 1 and 306 DF,  p-value: 5.344e-05
```

```r
summary(lm.lag3)
```

```
##
## Call:
## lm(formula = log(Cases) ~ Rain.lag3, data = mal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.11411 -0.48978 -0.04558  0.56805  1.90081
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.25582    0.09353  45.502   <2e-16 ***
## Rain.lag3    0.02979    0.01508   1.976    0.049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7585 on 305 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.01264,    Adjusted R-squared:  0.009405
## F-statistic: 3.905 on 1 and 305 DF,  p-value: 0.04904
```

```
summary(lm.lag4)
```

```
##
## Call:
## lm(formula = log(Cases) ~ Rain.lag4, data = mal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.05699 -0.49388 -0.01158  0.54307  2.15010
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.55766    0.09380  48.588   <2e-16 ***
## Rain.lag4   -0.02445    0.01510  -1.619    0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7591 on 304 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.008552,   Adjusted R-squared:  0.005291
## F-statistic: 2.622 on 1 and 304 DF,  p-value: 0.1064
```

Exercise 2

1. Plot the minimum temperature variable (`minT`) against time. What patterns do you observe?
2. Write down the equation of a linear regression model that accounts for a seasonal trend with a period of year. Fit the model defined in the previous point and add the curve of predicted values to the plot generate in the first point.
3. Add another sinusoidal function with period of 6 months to model fitted in the previous point. Generate the following plots: 1) one that displays both the seasonal trends estimated from the first model with a single sinusoidal functions with a one-year period, and the second model that add the sinusoidal curve with a six-month period; 2) the autocorrelgram based on the residuals from the two models. Based on these plots, which model do you believe to the best one?
4. Repear points 1 to 3, for `maxT` and `Rain`.