Exercise Day 3

Load the river blindness data, `Kericho.csv`. The data-set contains monthly reported malaria cases in Kericho, Kenya. The file `Kericho_explain.csv` provides a description of each of column in the data.

1. Generate a plot for the time series of the reported cases. What are the main features of the times series in terms of trend and seasonality?
2. Fit a linear model to the logarithm of the malaria cases with the following characteristics: 1) an linearly increasing trend over time with a change in slope on December 1983 (row no. 60 in the data) and a jump on February 1998 (row no. 230 in the data); 2) a seasonal trend with a period of 12 month.
3. Check if the residuals of the model fitted in the previous point show evidence of temporal correlation. Would adding an additional sinusoidal trend with a period of 6 months improve the fit of the model?
4. Based on the answer to the previous question, estimate an autoregressive process of the first order. Does this model explain all the residual temporal correlation?
5. Fit a Gaussian process model and use this to generate 95% confidence intervals for the reported malaria cases. Are the predictions from this model substantially different from those of the AR(1)?