# Some guidelines for Homework#3

1. Plot the series and briefly comment on the characteristics you observe (stationarity, trend, seasonality, ...).

2. Obtain a plot of the decomposition of the series, using stl(). Use an additive decomposition or a multiplicative one, depending on your data. Use the function forecast() to forecast future values. Multiplicative decomposition is useful when the logarithmic transformation makes a difference in the appearance of your series and shows more constant variance. Does the *remainder* look like a white noise to you? White noise is just a group of independent, identically distributed variables, with zero mean and constant variance. Answer to this point just visually or plot the ACF and PACF of the *remainder* part.

3. Fit an ARIMA model to your time series. Some steps to follow:

   *a*) Decide on whether to work with your original variable or with the log transform one (use plot and tsdisplay of both variables, the transformed one and the original one). Remember that logarithmic transformation is useful to stabilize the variance. If both plots look almost the same to you, just use the original data.

   *b*) Are you going to consider a seasonal component? If the answer is yes, identify $s$. You can use the periodogram (function tsdisplay with parameter "plot.type=spectrum"), ACF (with significant correlations on lag $s$ and its multiples), seasonplot() and monthplot(). For montlhy data, $s = 12$; for quarterly data $s = 4$. Declare this frequency when you define your time-series in R, regardless of including or not a seasonal term in your model. This will permit you to use the functions *seasonplot (package forecast)* or *monthplot (built-in function)*.

   *c*) Decide on the values of $d$ and $D$ to make your series stationary. $D$ values are not usually greater than 1. You can use the standard deviation procedure and stationary tests (adf.test, kpss.test). Also, functions like *ndiffs* and *nsdiffs* may be useful. Don't be surprised if that group of tools give you contradictory results. Just, make decision to keep going. Plots of the ACF and PACF also help. If you can draw a pattern out of the ACF and PACF function, then stop differencing and start modelling. If that pattern doesn't work, then consider rethinking the part where you took differences.

   *d*) Identify values for $p$ and $q$ for the regular part and $P$ and $Q$ for the seasonal part. Start with low values and then increase them, one at a time. Fit different models and compare them using AICc and checking the residuals. Check also the correlation between the coefficients of the model.

   *e*) Make diagnostic of the residuals for the final model chosen (autocorrelations, zero mean, normality). Use plots and tests.

   *f*) Once you have found a suitable model, repeating the fitting model process several times if necessary, use it to make forecasts. Plot them.

   *g*) Use the function `getrmse` to compute the test set RMSE of some of the models you have already fiited. Which is the one minimizing it? Use the last year of observations (12 observations for monthly data, 4 observations for quarterly data) as the test set.

*h*) You can also use the auto.arima() function with some of its parameters fixed, to see if it suggests a better model that the one you have found. Don't trust blindly its output. Automatic found models aren't based on an analysis of residuals but in comparing some other measures like AIC. Depending on how complex the data set is, they may find models with high values for $p, q, P$ or $Q$ (greater than 2).