Lab 5

Pstat 174/274

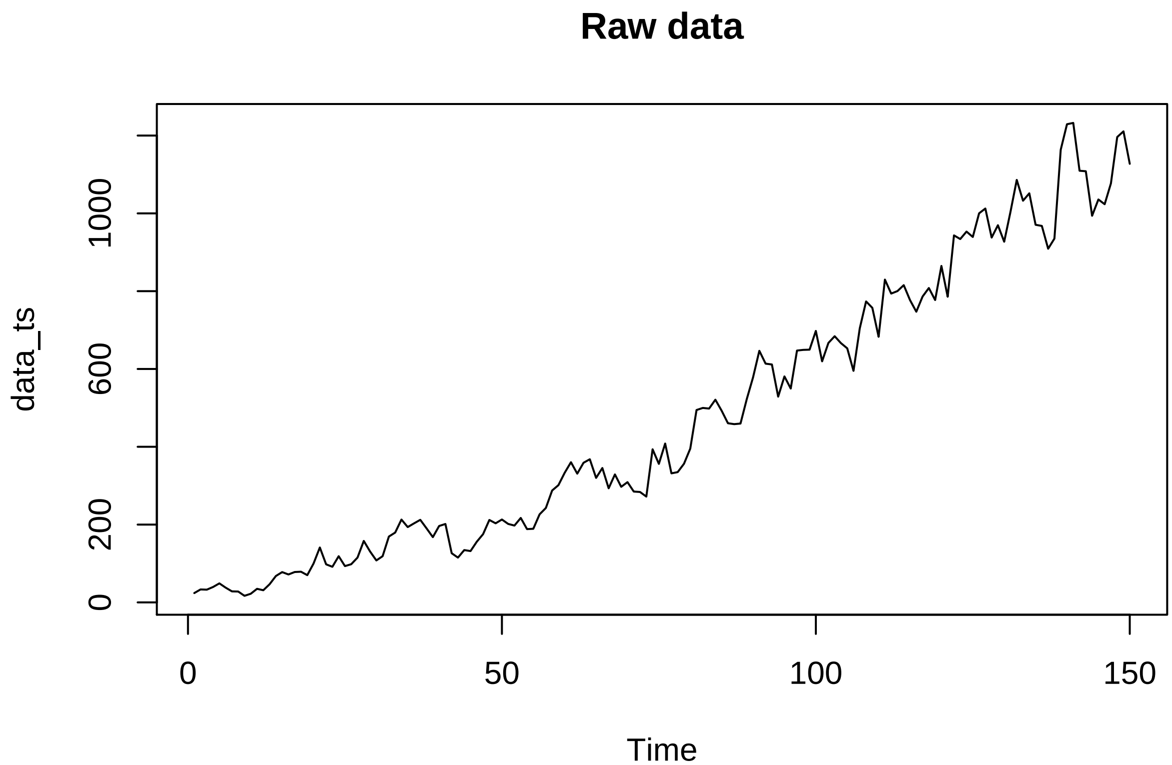
January 31, 2017

a. Download the file `lab5data.txt` and import the data into `R` using `scan`.

```
# Set working directory
setwd("/Users/jzapata/Dropbox/Winter 17/TA 174-274/Labs")
# Read data
?scan
data <- scan("lab5data.txt")
```

b. Plot the time series.

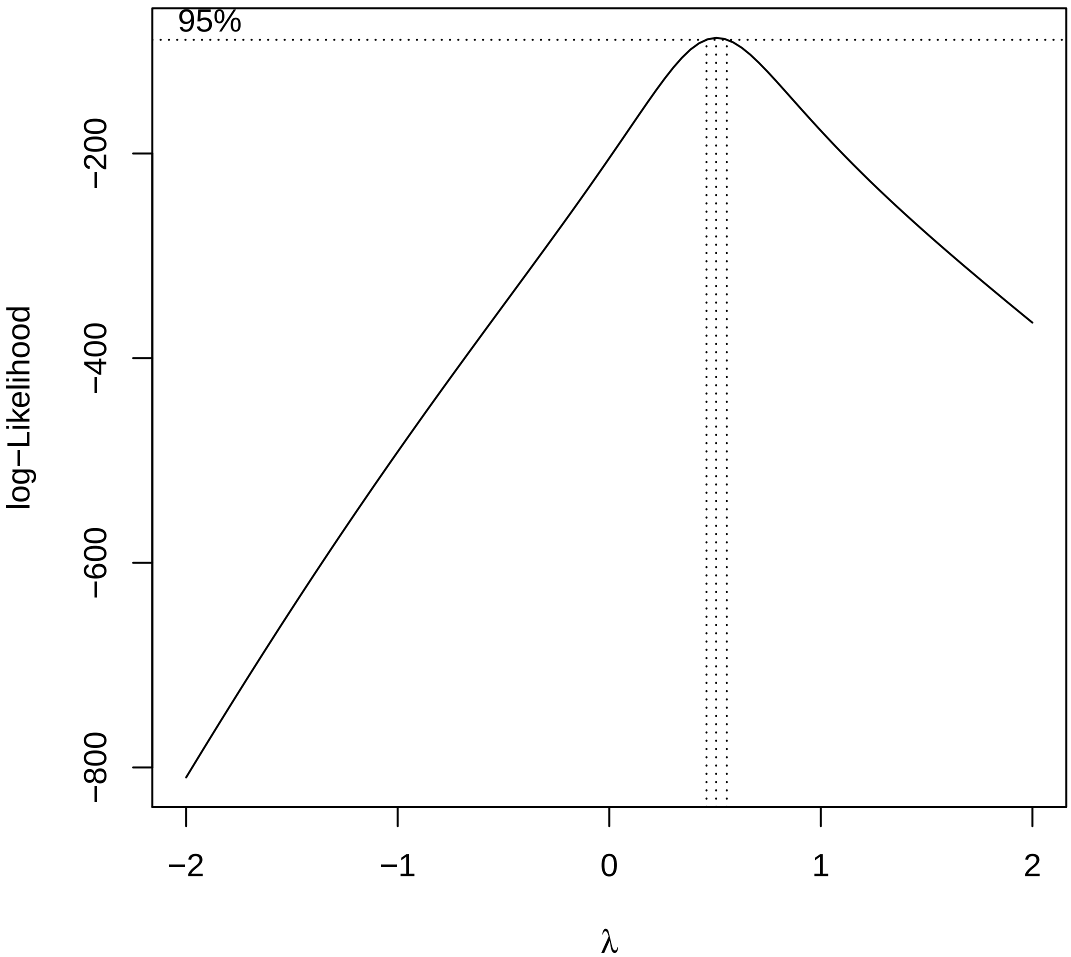
```
# Convert to time series object
data_ts <- ts(data)
ts.plot(data_ts,main  = "Raw data")
```



c. Is the variance constant over time? Is a transformation necessary? Try square root, log, and boxcox transformations to stabilize the variance.

```
# Transform data using boxcox()
require(MASS)
bcTransform <- boxcox(data_ts ~ as.numeric(1:length(data_ts)))
lambda = bcTransform$x[which(bcTransform$y == max(bcTransform$y))]
lambda
```

```
## [1] 0.5050505
```



Notice that the 95% confidence interval for λ includes the value $\lambda = 1/2$ (in fact the estimated MLE of the Box-Cox tranformation is $\hat{\lambda} \approx 1/2$). Therefore, we implement the Box-Cox tranformation as:

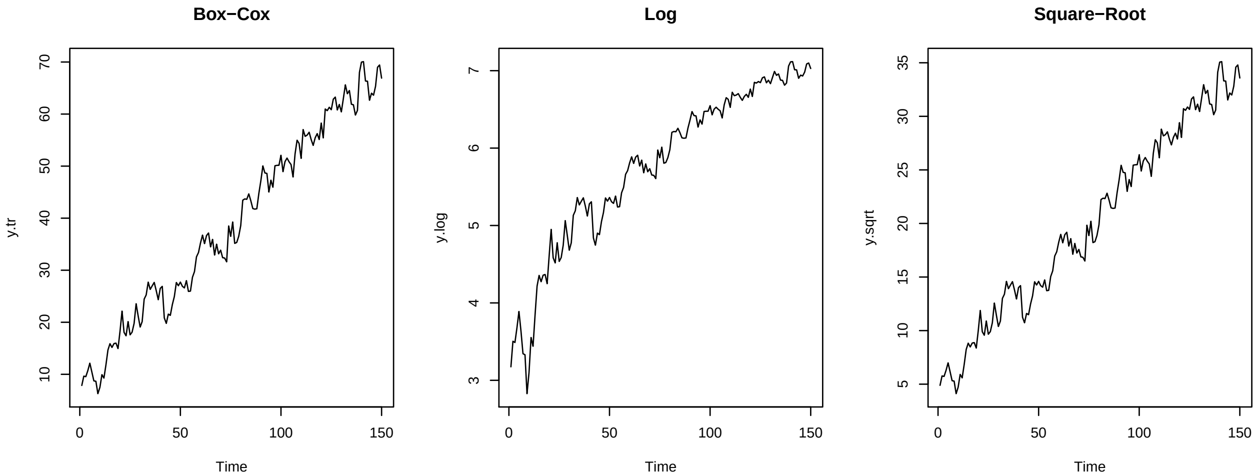
```
y.tr <- (1/lambda)*(data_ts^lambda - 1) # Box-cox
```

as well as the other tranformations:

```
y.log <- log(data_ts) # log
y.sqrt <- sqrt(data_ts) # sqrt
```

We now plot the transformed time-series:

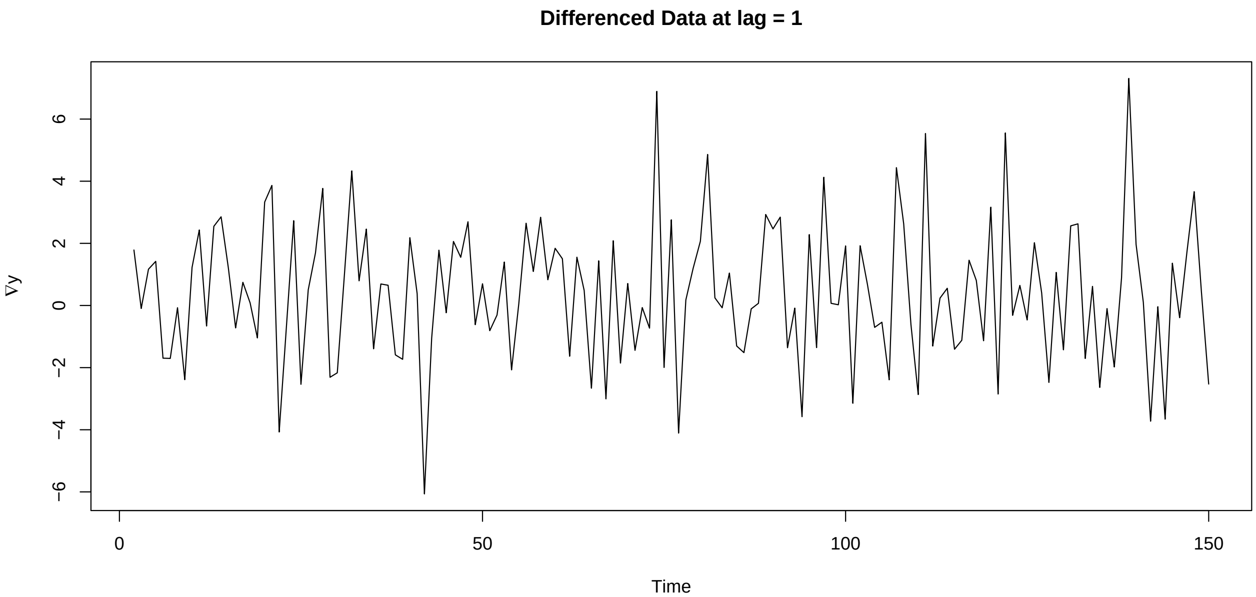
```
op <- par(mfrow = c(1,3))
ts.plot(y.tr,main = "Box-Cox")
ts.plot(y.log,main = "Log")
ts.plot(y.sqrt,main = "Square-Root")
par(op)
```



d. Is their a trend in the data? linear? quadratic? seasonal? Try differencing the data to remove the trend.

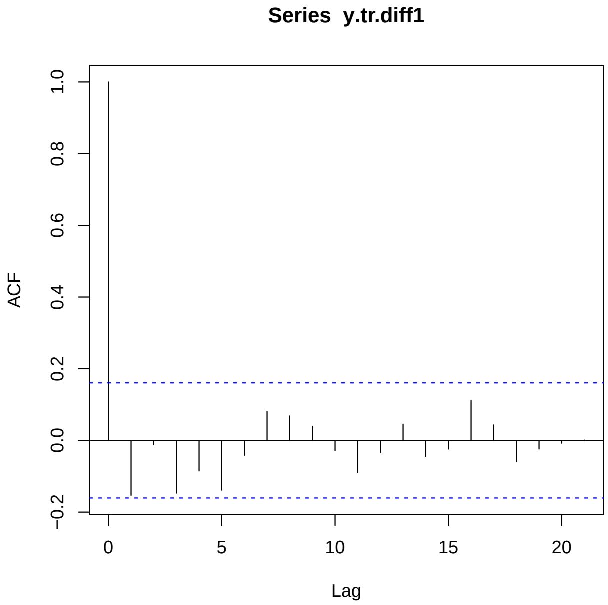
There is a strong linear trend in either of the transformed series (e.g. Box-Cox). Therefore, in order to remove the trend component it is enough to difference the data at lag 1, i.e.:

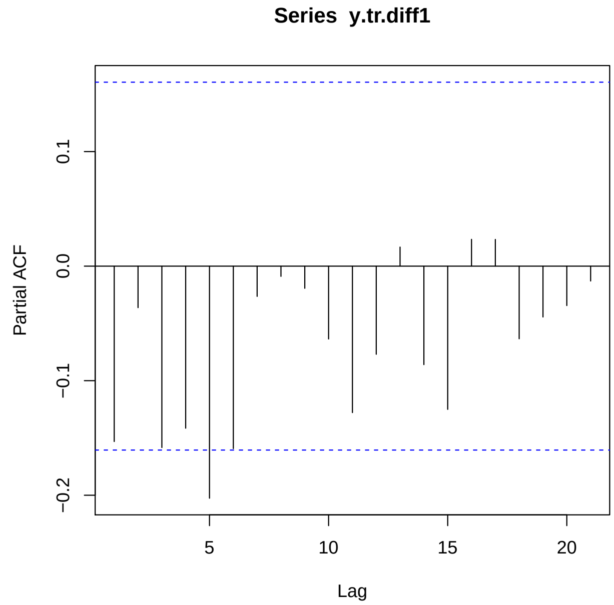
```
y.tr.diff1 <- diff(y.tr,1)
ts.plot(y.tr.diff1,main = "Differenced Data at lag = 1",ylab=expression(paste(nabla,y)))
```



e. Plot the ACF and PACF. What model do you think generated the data?

```
op <- par(mfrow = c(1,2))
acf(y.tr.diff1)
pacf(y.tr.diff1)
par(op)
```





- ACF: Since all the sample ACF values lie within the 95% confidence bounds we conclude that no significant MA component is present (since otherwise we would have significant ACF values at lags below the order q of the MA process). Morover, notice the oscillating and decaying behavior of the ACF values indicate the presence of an AR process – which we now confirm via the PACF.
- PACF: There is a significant sample PACF at lag 5. Moreover, after lag 5 all sample PACF are not significant (lie within the confidence bounds) which match the theoretical behaviour of the PACF of an AR(5) model.

f. If you suspect the data follows an AR model, estimate the coefficients using Yule-Walker estimation (Type `help(ar)` for reference on how to estimate AR model coefficients using Yule-Walker estimation).

```r
# Fit ar model using yule-walker equations
?ar
(fit <- ar(y.tr.diff1, method="yule-walker"))
```

```
##
## Call:
## ar(x = y.tr.diff1, method = "yule-walker")
##
## Coefficients:
##       1        2        3        4        5        6
## -0.2478  -0.1354  -0.2270  -0.1964  -0.2369  -0.1593
##
## Order selected 6  sigma^2 estimated as  4.523
```

Notice that `ar` automatically determines the order of an the AR model via AIC.