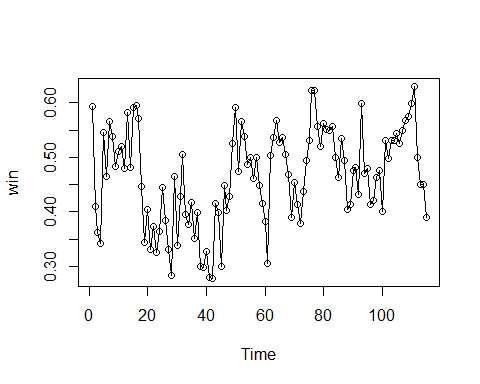
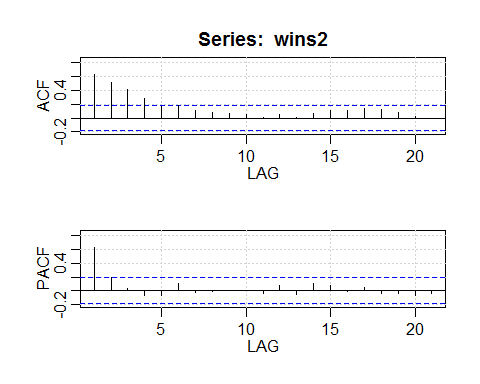
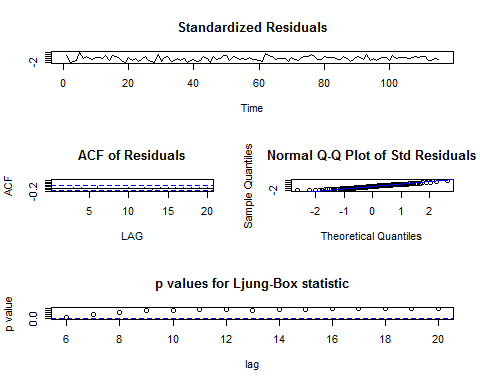
3) A time series plot of the Phillies win percentage from 1901-2015 is shown in the plot below, along with ACF and PACF plots. The time series plot doesn't indicate much of a long term trend, so no differencing seems necessary.





The ACF plot indicates a 4 term moving average would be good to use since after a lag of 4 the time series appears like white noise, and the PACF plot indicates 1 AR term may be good to use for the same reason. Fitting this ARMA(1,4) model gives the following results:

model.1 <- sarima(wins2, 1, 0, 4, details=F)



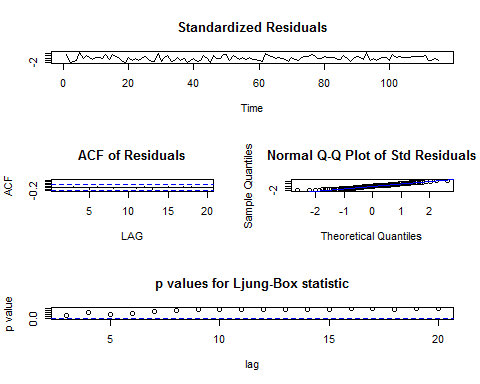
model.1  
#   
## Coefficients:  
## ar1 ma1 ma2 ma3 ma4 xmean  
## 0.5040 0.0268 0.2212 0.1864 0.0402 0.4664  
## s.e. 0.8175 0.8621 0.4194 0.3785 0.3623 0.0177  
##   
## sigma^2 estimated as 0.004195: log likelihood = 151.23, aic = -288.45  
##   
## $AIC  
## [1] -4.369499  
##   
## $AICc  
## [1] -4.343006  
##   
## $BIC  
## [1] -5.226285

The standardized residuals and the ACF of the residuals look like white noise, the normal Q-Q plot indicates the residuals are normally distributed and the p-values are all very high. The aic value is -288.45. To see if the number of terms can be reduced while still maintaining a low aic value, I created a matrix using the code from the class website which indicates I could lower the aic value by including fewer terms.

## [,1] [,2] [,3] [,4] [,5]  
## [1,] -232.1977 -267.6920 -278.0910 -287.4531 -290.3753  
## [2,] -291.3912 -292.6096 -291.1951 -290.4367 -288.4520  
## [3,] **-292.8598** -290.8842 -289.6750 -288.4382 -288.0382  
## [4,] -290.9112 -289.0216 -289.3576 -288.3199 -287.1955  
## [5,] -289.8709 -287.4229 -291.8545 -287.3107 -284.7335

It seems like a model with 2 auto-regressive terms and no moving average terms would produce the highest aic value.

model.2 <- sarima(wins2, 2, 0, 0, details=F)



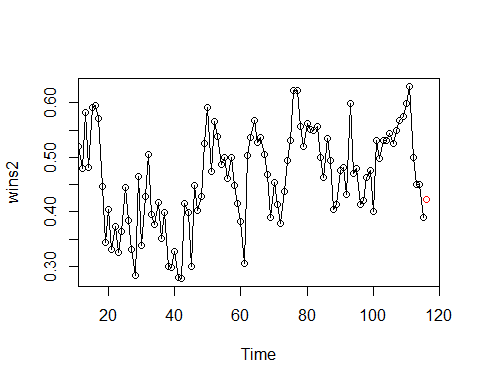
model.2  
#  
## Coefficients:  
## ar1 ar2 xmean  
## 0.5295 0.1768 0.4658  
## s.e. 0.0941 0.0941 0.0202  
##   
## sigma^2 estimated as 0.004257: log likelihood = 150.43, aic = -292.86  
##   
## $AIC  
## [1] -4.407071  
##   
## $AICc  
## [1] -4.386518  
##   
## $BIC  
## [1] -5.335464

This is my preferred model since it has fewer terms, a lower aic value and all of the diagnostic plots look the same or better. Looking at the estimates relative to the standard errors, the 2 coefficients seem significant. The only diagnostic plot that I think could be improved is the normal Q-Q plot, which currently shows some skewness, especially in the lower tail.

The full equation for the model is Xt = 0.5295*X*t-1 + 0.1768*X*t-2 + *W*t.

Using the model to predict the Phillies win percentage in 2016 indicates the win percentage will be 0.422 in 2016.

sarima.for(wins2, 1, 2, 0, 0, 0, 0, 0, 1)



## $pred  
## Time Series:  
## Start = 116   
## End = 116   
## Frequency = 1   
## [1] 0.4225115  
##   
## $se  
## Time Series:  
## Start = 116   
## End = 116   
## Frequency = 1   
## [1] 0.06524392