

M445 HW #9

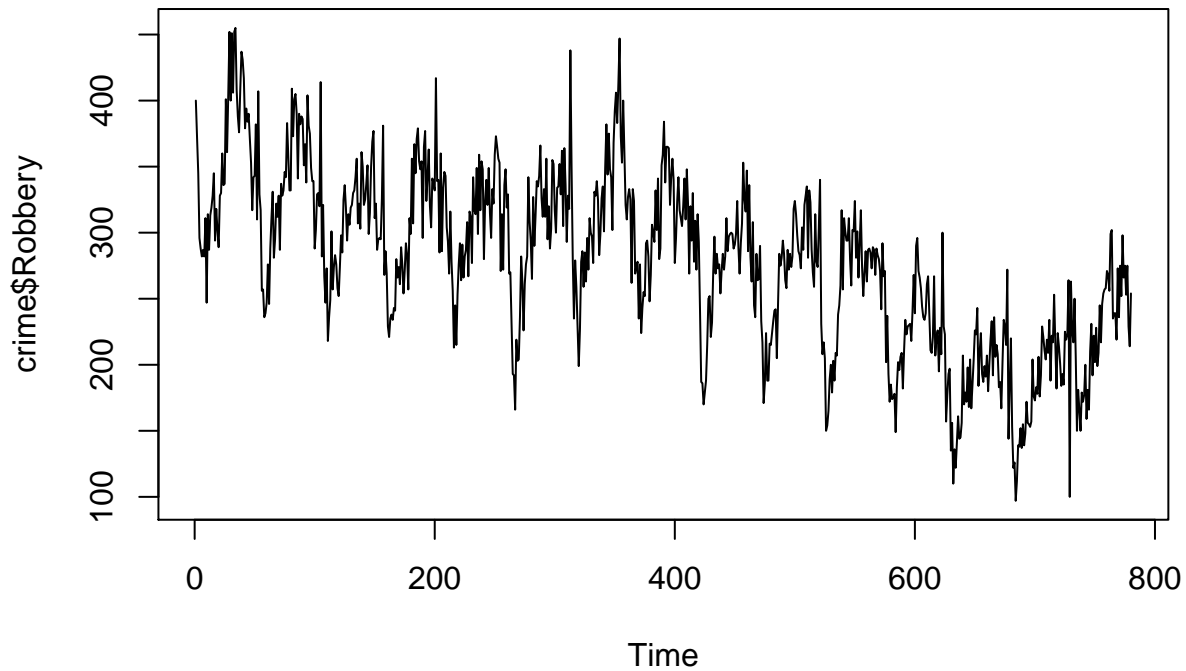
Erick Castillo

5/7/2021

The data set used for these sets of problems is a time series regarding the amount of crimes in Chicago.

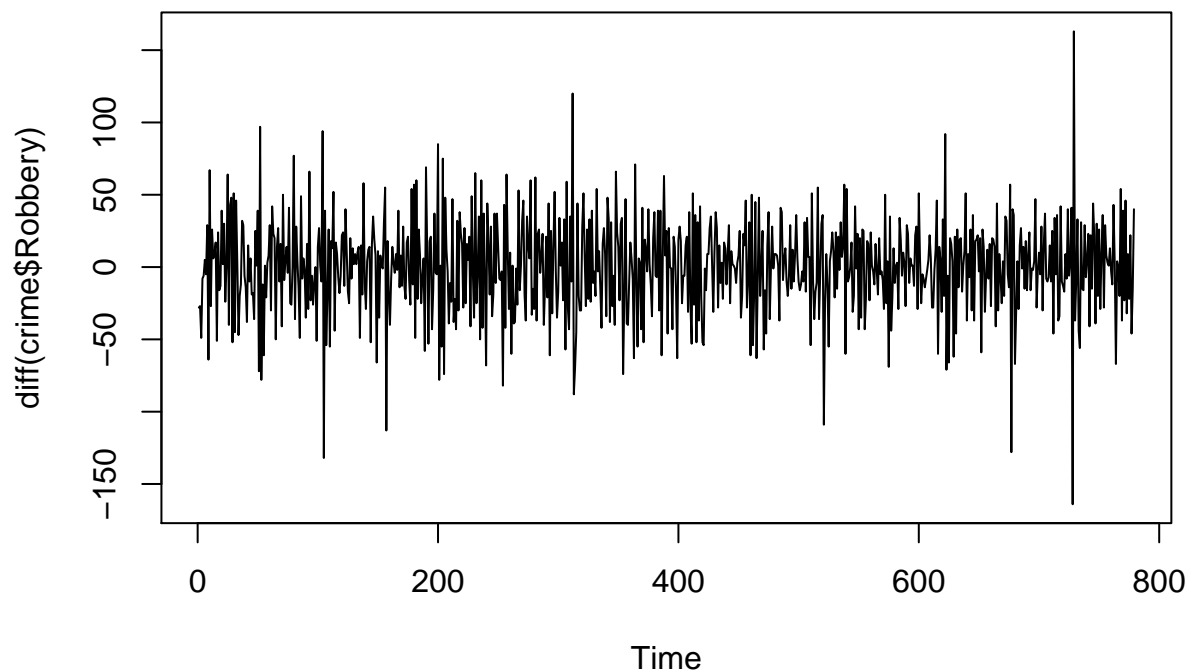
A. Plot the time series data. Does the data appear to be stationary? If not, modify the data so that it is, then plot it.

```
plot.ts(crime$Robbery)
```



The data does not appear to be stationary. It seems to be trending downwards as the weeks pass. This can be solved by differencing the data. A nifty function in R, `diff()`, does just this.

```
plot.ts(diff(crime$Robbery))
```



```
adf.test(diff(crime$Robbery))
```

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag   ADF p.value
## [1,]  0 -43.3   0.01
## [2,]  1 -26.7   0.01
## [3,]  2 -20.9   0.01
## [4,]  3 -17.4   0.01
## [5,]  4 -14.9   0.01
## [6,]  5 -13.4   0.01
## [7,]  6 -12.9   0.01
## Type 2: with drift no trend
##      lag   ADF p.value
## [1,]  0 -43.3   0.01
## [2,]  1 -26.6   0.01
## [3,]  2 -20.9   0.01
## [4,]  3 -17.4   0.01
## [5,]  4 -14.9   0.01
## [6,]  5 -13.4   0.01
## [7,]  6 -12.9   0.01
## Type 3: with drift and trend
##      lag   ADF p.value
```

```
## [1,] 0 -43.3 0.01
## [2,] 1 -26.6 0.01
## [3,] 2 -20.9 0.01
## [4,] 3 -17.4 0.01
## [5,] 4 -14.9 0.01
## [6,] 5 -13.4 0.01
## [7,] 6 -12.8 0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

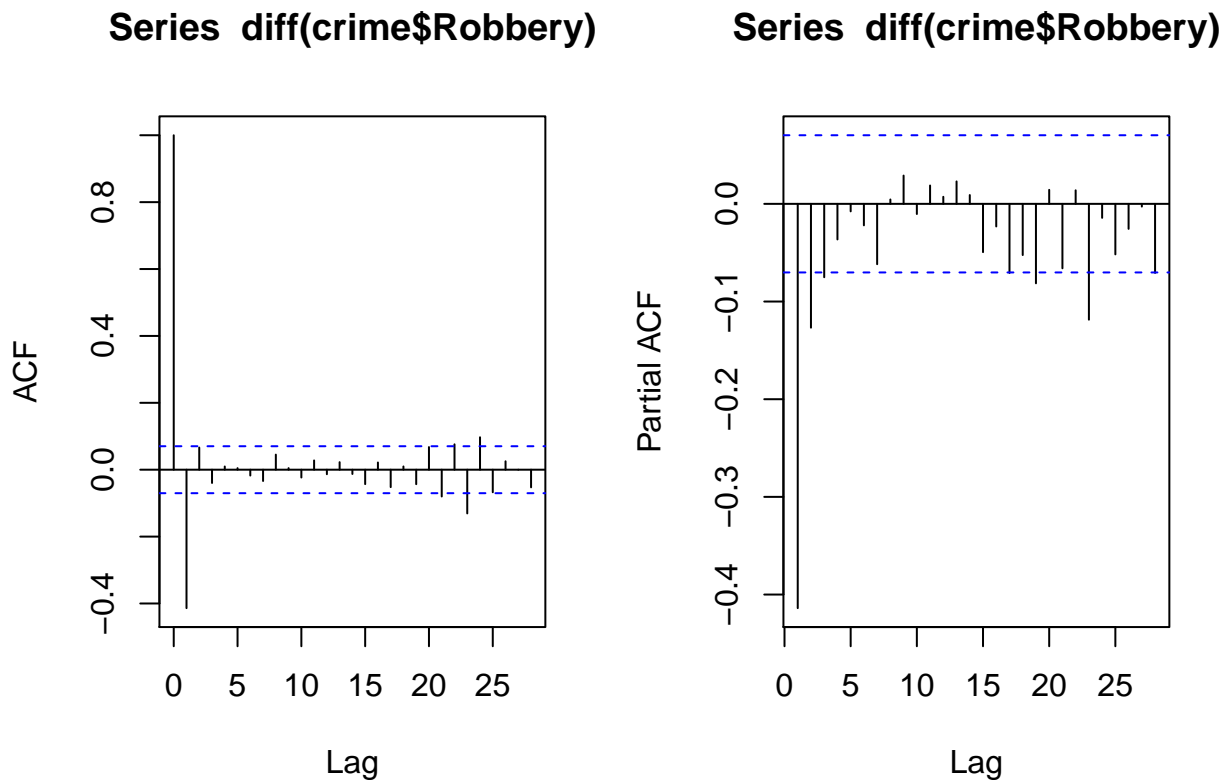
Notice this new plot appears to be much more stationary than the original plot. The p-values from the ACF test also indicate that the differenced data is stationary.

B. Is there a visual seasonality trend? If there is, then what model should be used to fit the data?

The data does appear to have a seasonality trend. The original plot, the one without differencing follows a decreasing sine function pattern, with peaks and troughs which come at consistent intervals. Because of this seasonality, I would propose using either an autoregressive model $AR(p)$ or $MA(q)$.

C. Plot the ACF and PACF of the differenced time series. After studying the plots, what model should be used?

```
par(mfrow = c(1,2))
acf(diff(crime$Robbery))
pacf(diff(crime$Robbery))
```



The left plot corresponds to the ACF, the right to the PACF. Viewing these plots, we can see that the model that would best fit the data would either be an AR(2) model or an MA(1) model with a seasonality effect of 23.

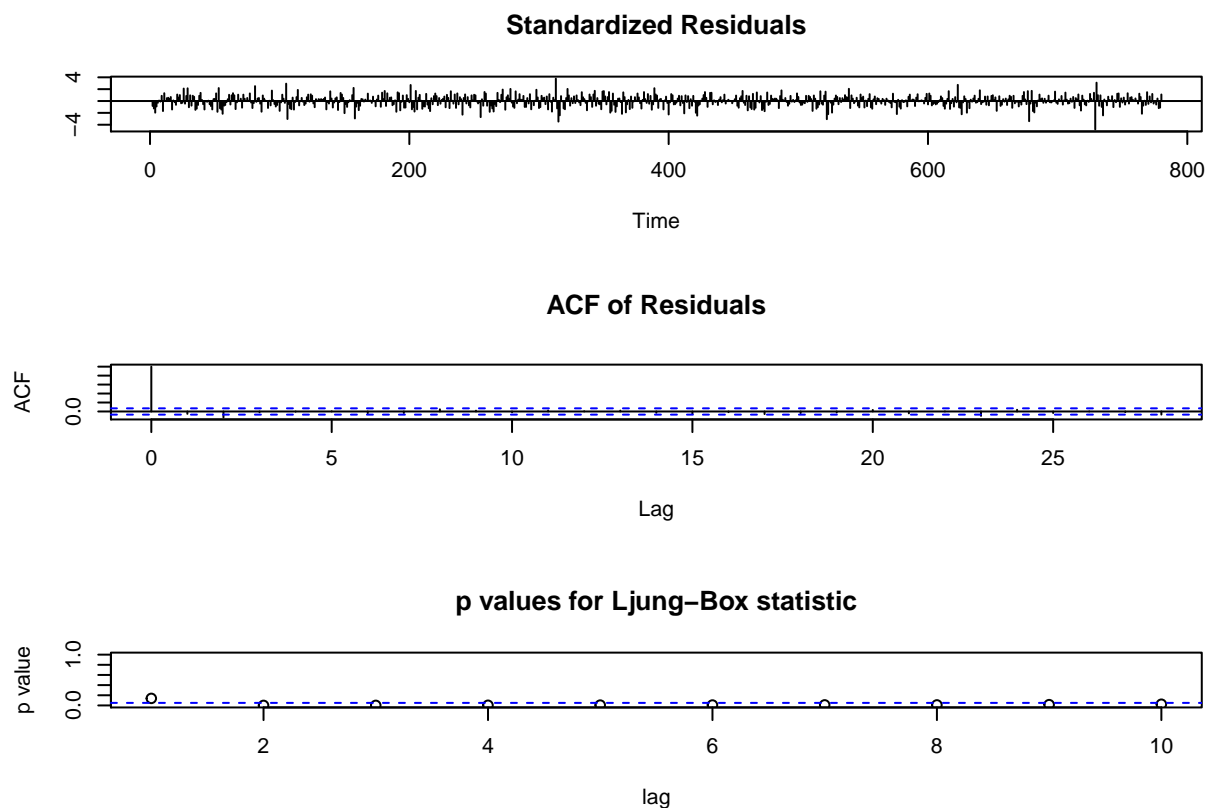
D. Fit an AR(1) model. Report the coefficients and use tsdiag to obtain the diagnostics plots.

```
ts.mod1 <- arima(crime$Robbery, order = c(1,1,0))
ts.mod1

##
## Call:
## arima(x = crime$Robbery, order = c(1, 1, 0))
##
## Coefficients:
##          ar1
##        -0.4145
## s.e.      0.0326
##
## sigma^2 estimated as 941.5:  log likelihood = -3772.55,  aic = 7549.1
```

The AR(1) model takes the form $X_t = \alpha X_{t-1} + \epsilon_t$. In this case, $\alpha = -0.4145$.

```
tsdiag(ts.mod1)
```



The first two diagnostic plots look great. The last one does not. This means that there is some autocorrelation still present in the model. The model must be corrected.

E. Fit a few models and compare their AIC's. Base these models of prior results. Which is the better model?

```
ts.mod2 <- arima(crime$Robbery, order = c(1,1,0), seasonal = list(order = c(1,1,0), period = 23))
ts.mod3 <- arima(crime$Robbery, order = c(0,1,1), seasonal = list(order = c(0,1,1), period = 23))
ts.mod4 <- arima(crime$Robbery, order = c(0,1,1))
ts.mod5 <- arima(crime$Robbery, order = c(0,1,1), seasonal = list(order = c(1,1,1), period = 23))
ts.mod6 <- arima(crime$Robbery, order = c(0,1,1), seasonal = list(order = c(0,1,1), period = 12))
```

```
AIC(ts.mod1)
```

```
## [1] 7549.104
```

```
AIC(ts.mod2)
```

```
## [1] 7703.672
```

```
AIC(ts.mod3)
```

```
## [1] 7402.285
```

```
AIC(ts.mod4)
```

```
## [1] 7532.734
```

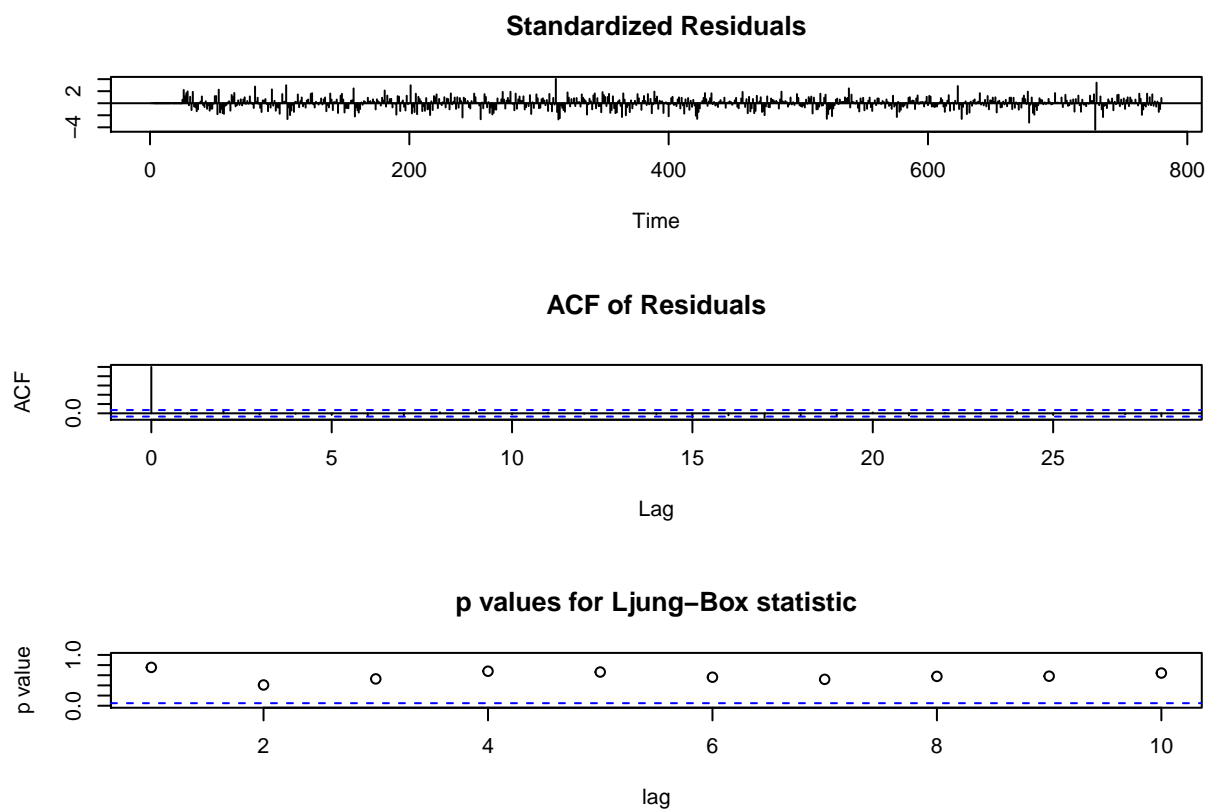
```
AIC(ts.mod5)
```

```
## [1] 7394.267
```

```
AIC(ts.mod6)
```

```
## [1] 7467.853
```

```
tsdiag(ts.mod5)
```



Notice from the above AIC outputs that my fifth model has the smallest AIC. This is an $I(1)MA(1)$ with a seasonality effect of period= 23 for each the $AR(1)$ and the $MA(1)$. This would be my final model based solely on AIC. The diagnostics plot for this model indicate that it is healthy and useful for predictions.

F. Compute and plot predictions going 52 weeks out into the future.

```
detach("package:aTSA", unload=TRUE)

pred <- forecast(ts.mod5, h = 52)
plot(pred)
```

Forecasts from ARIMA(0,1,1)(1,1,1)[23]

