Time Series Analysis of Traffic Collision in Los Angeles

Zheqi Wu
March 2018

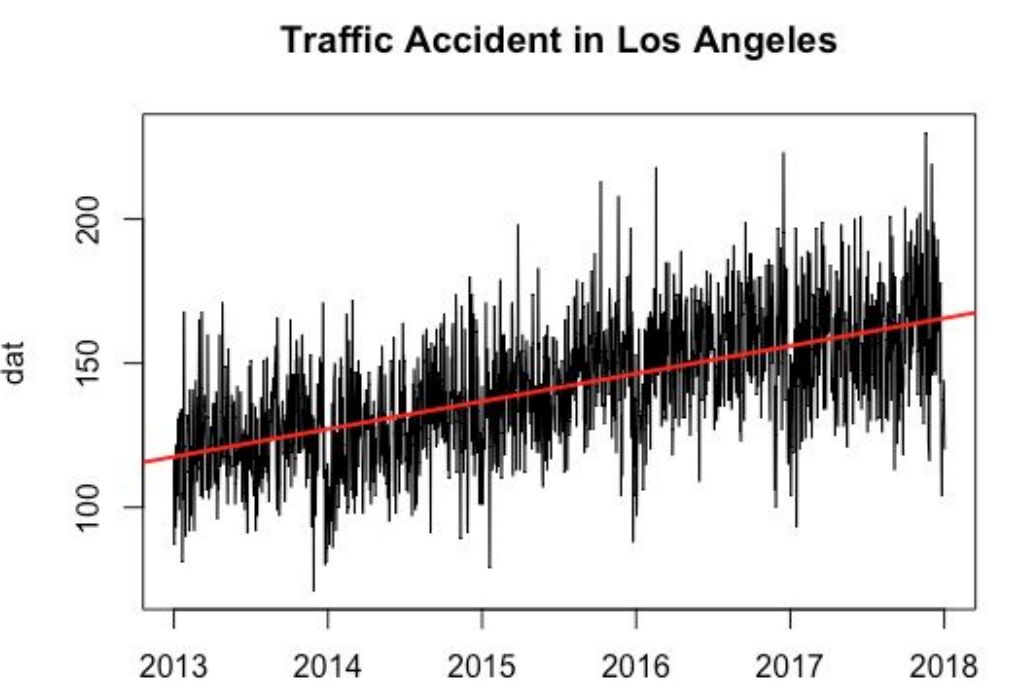
Data Description

The dataset is download from the DATA.GOV website which contains every reported traffic collision incident in the City of Los Angeles dating back to 2010.

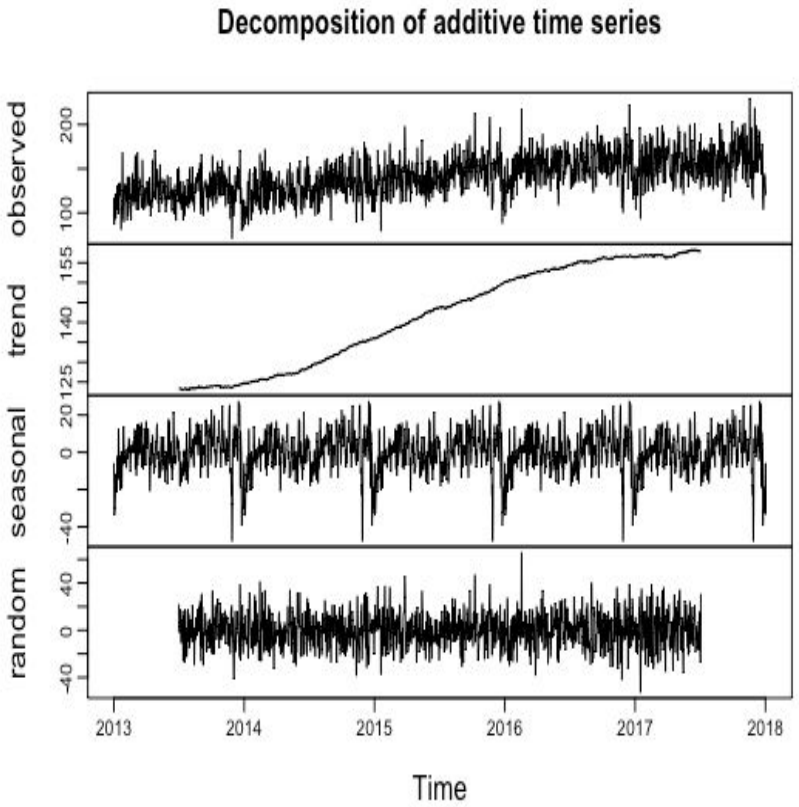
So I count the number of traffic incidents every day from 01/01/2013 to 12/31/2017 to construct my time series dataset with 1827 observations.

Hold daily data from 01/01/2018 to 01/31/2018 as the testing data.

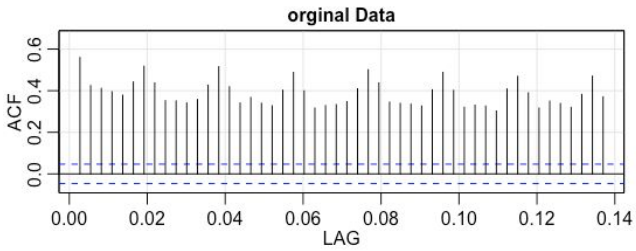
Data Exploration

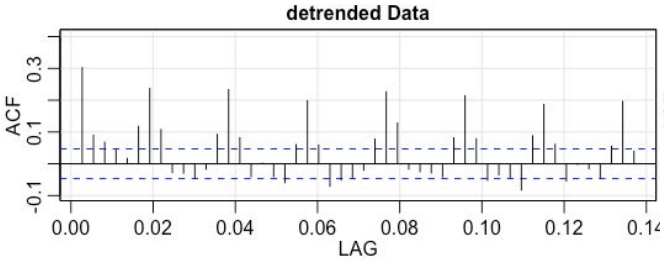


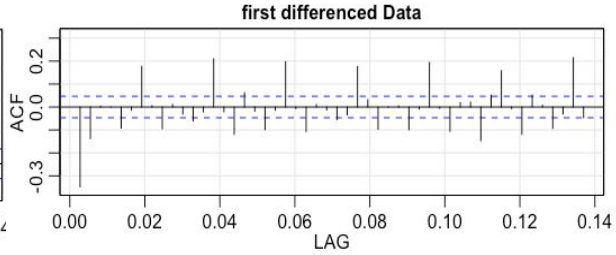

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.923e+04  5.940e+02  -32.37   <2e-16 ***
time(dat)    9.611e+00  2.947e-01   32.61   <2e-16 ***
```

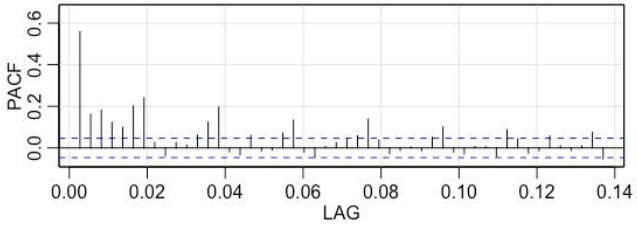


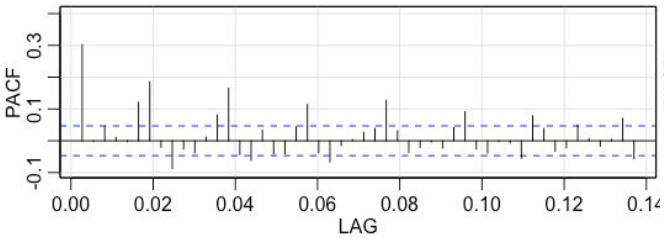
Data Exploration

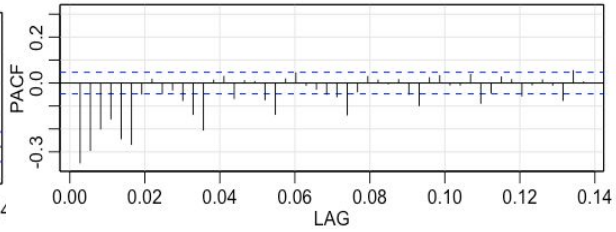




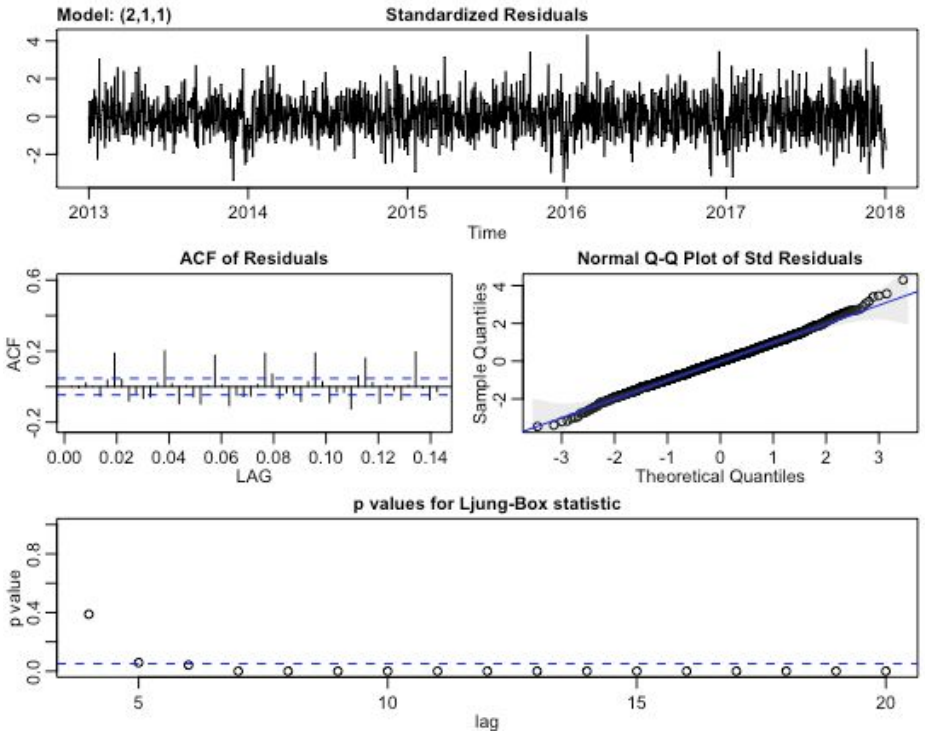






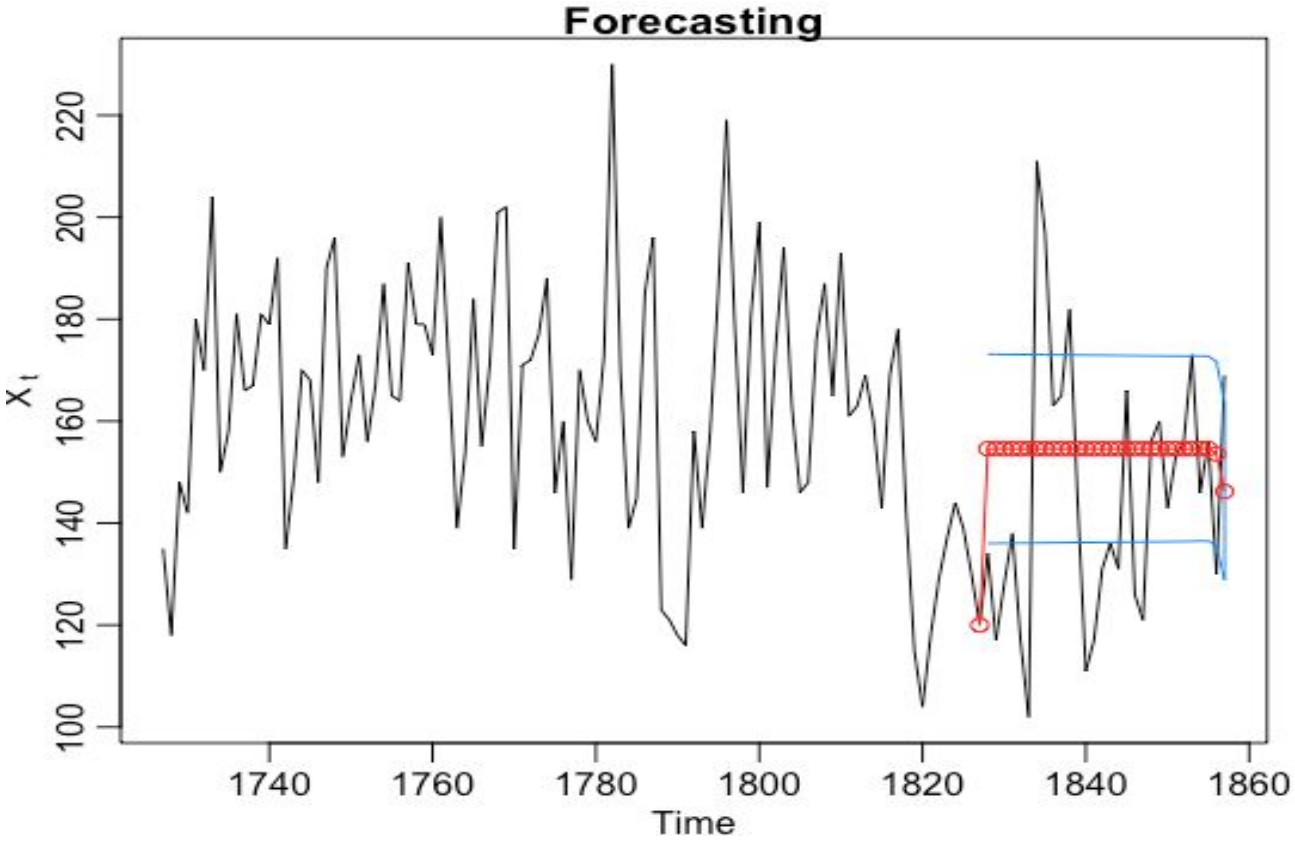


Model Fitting and Diagnostics

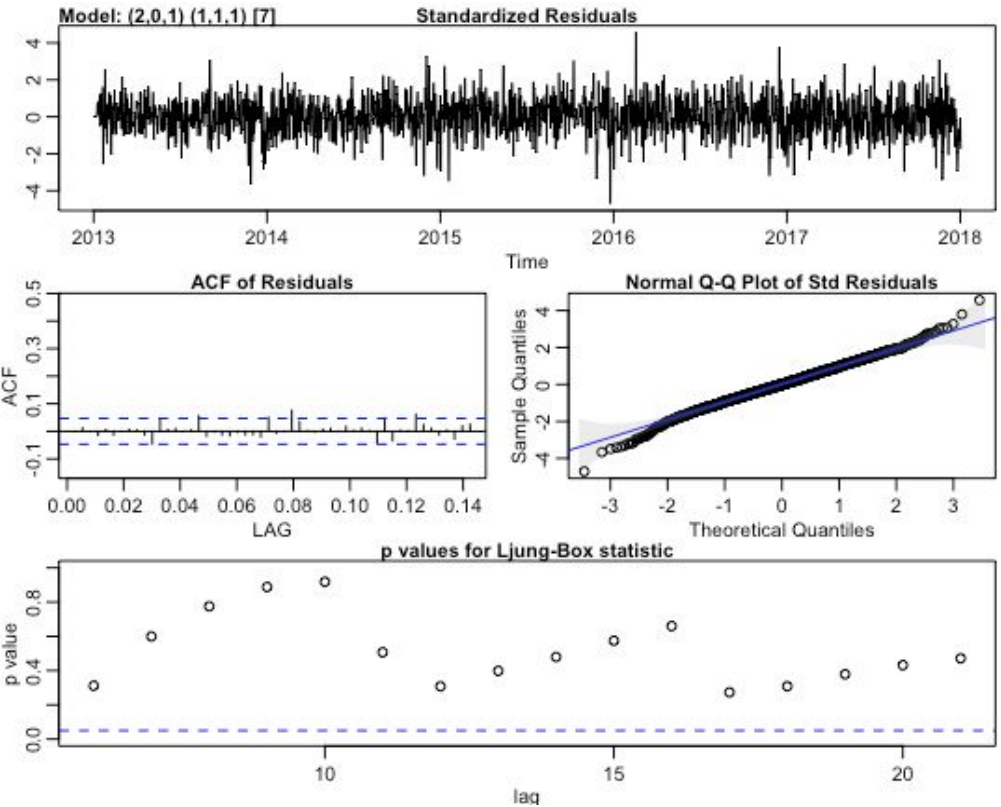


ARIMA model´	AIC	BIC
(2,1,1)	**6.706121**	**5.718186**
(2,1,2)	6.706919	5.721999
(2,0,1)	6.707929	5.719994
(2,0,2)	6.707674	5.722755
(1,0,2)	6.763829	5.772877

Model Fitting and Diagnostics

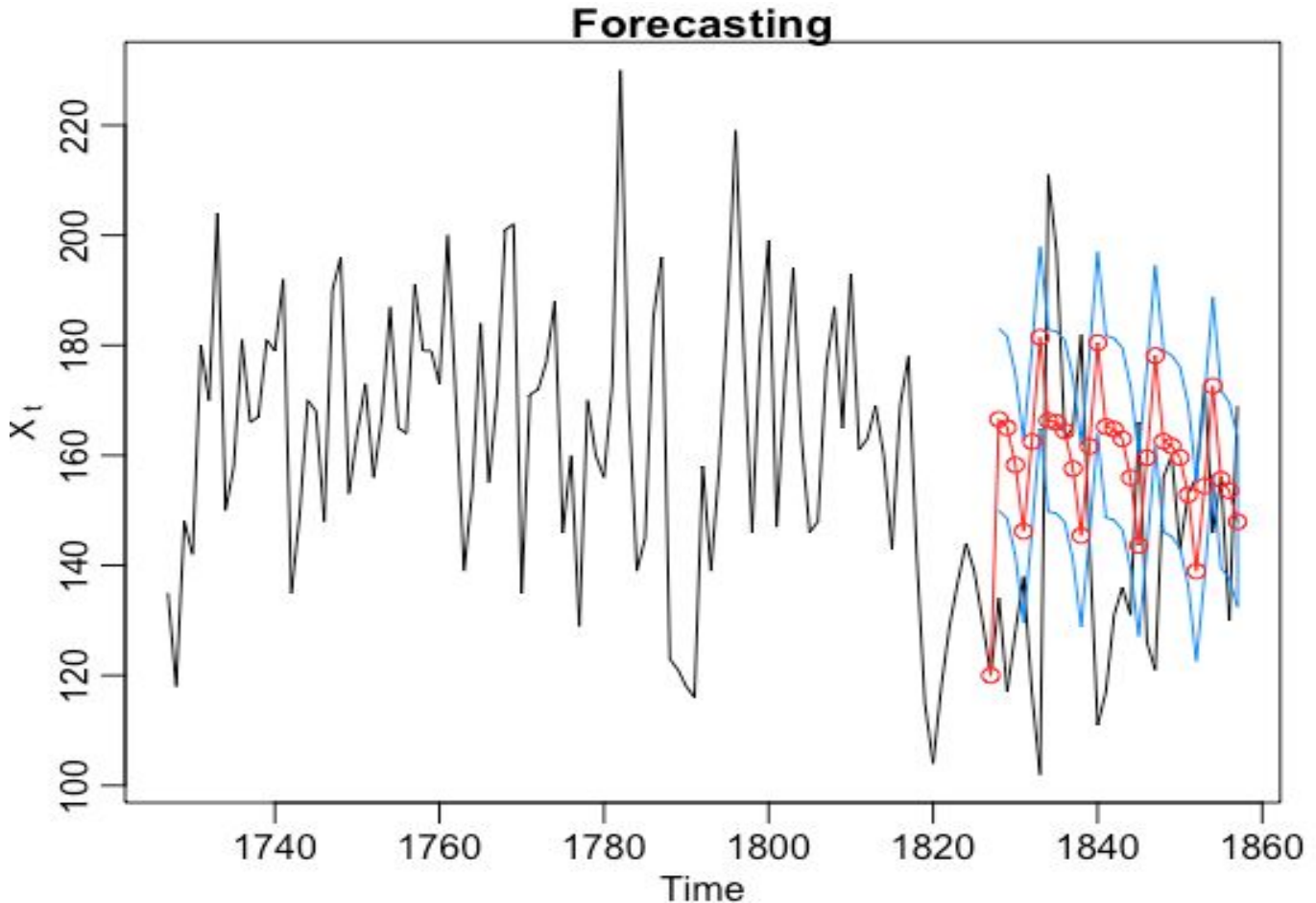


Model Fitting and Diagnostics

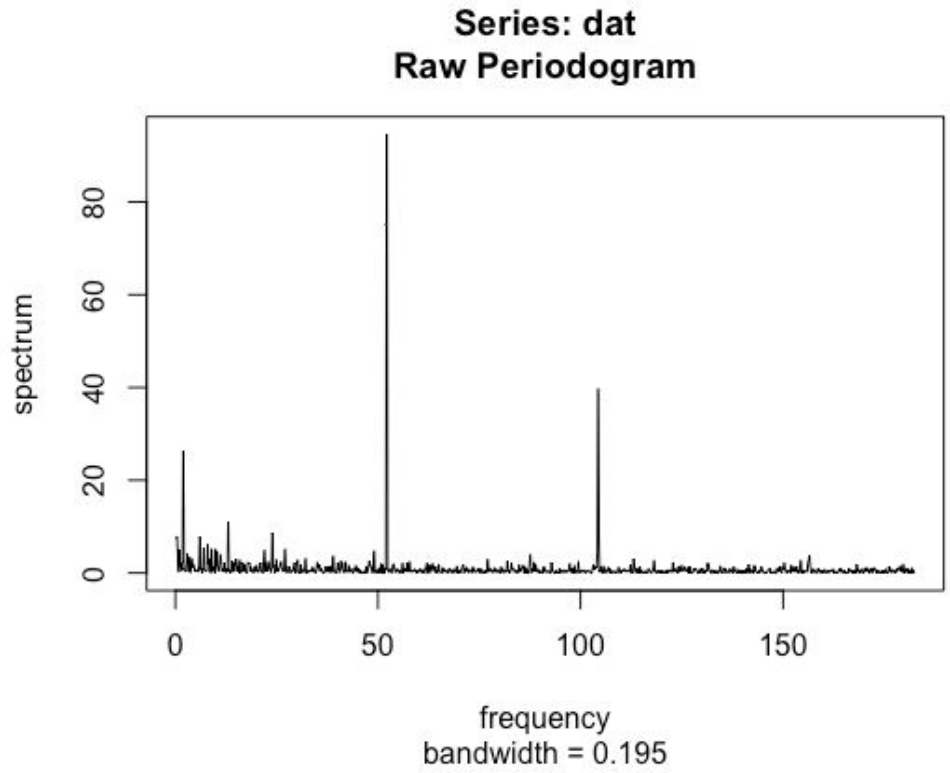


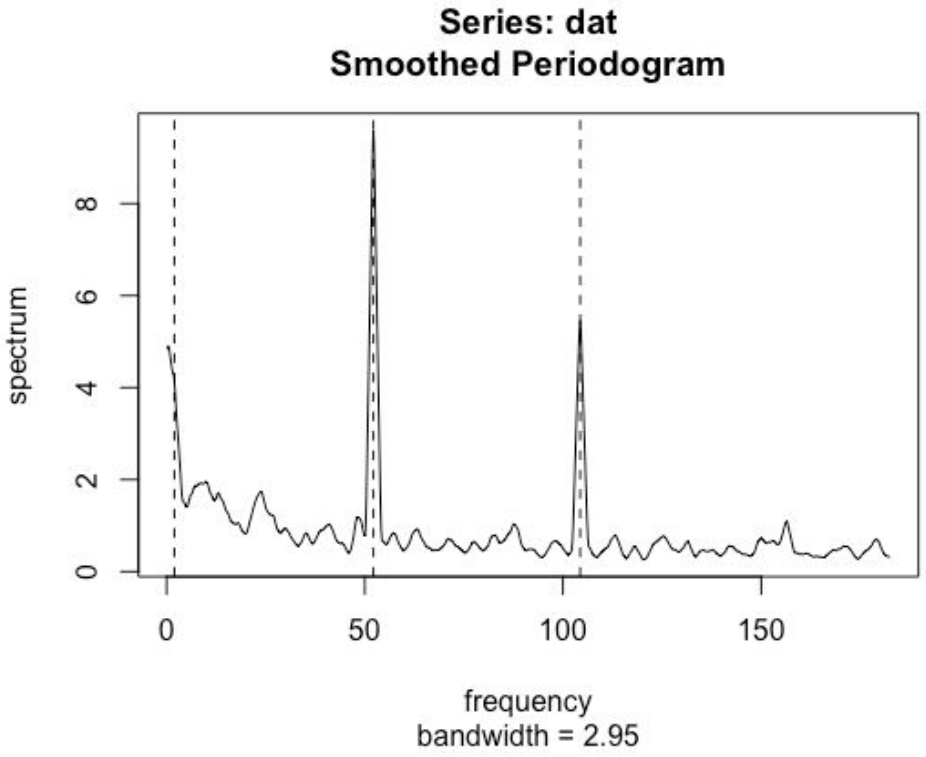
ARIMA model	AIC	BIC
(2,0,1)(1,1,1)[7]	**6.484181**	**5.49734**
(2,0,1)(0,1,1)[7]	6.484315	5.499396
(1,0,1)(0,1,1)[7]	6.485275	5.502278
(1,0,1)(1,1,1)[7]	6.486009	5.50109
(2,1,1)(0,1,1)[7]	6.492648	5.504713
(2,1,1)(1,1,1)[7]	6.493552	5.508633

Model Fitting and Diagnostics

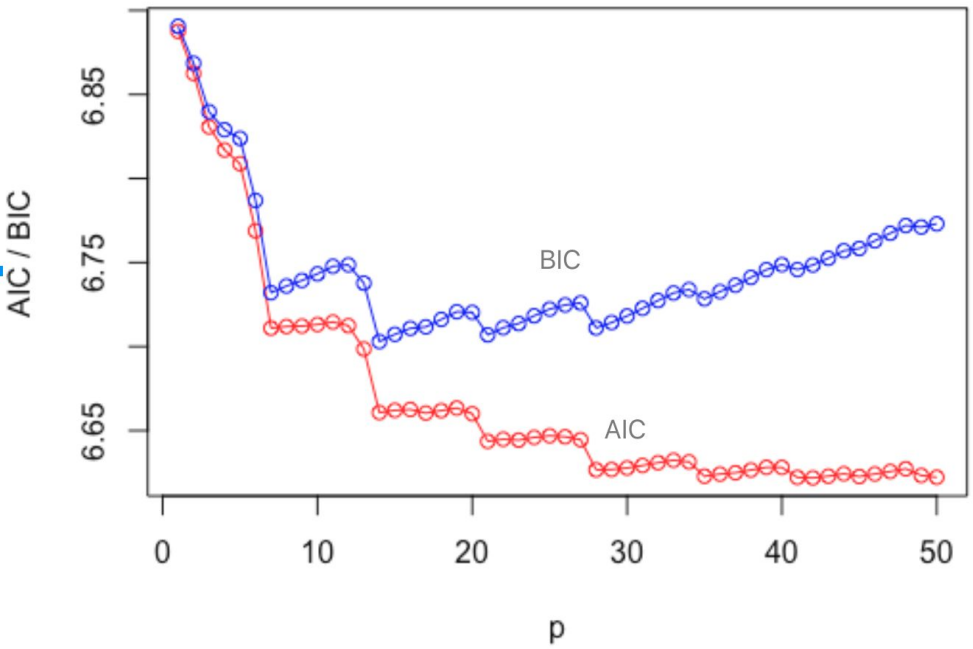


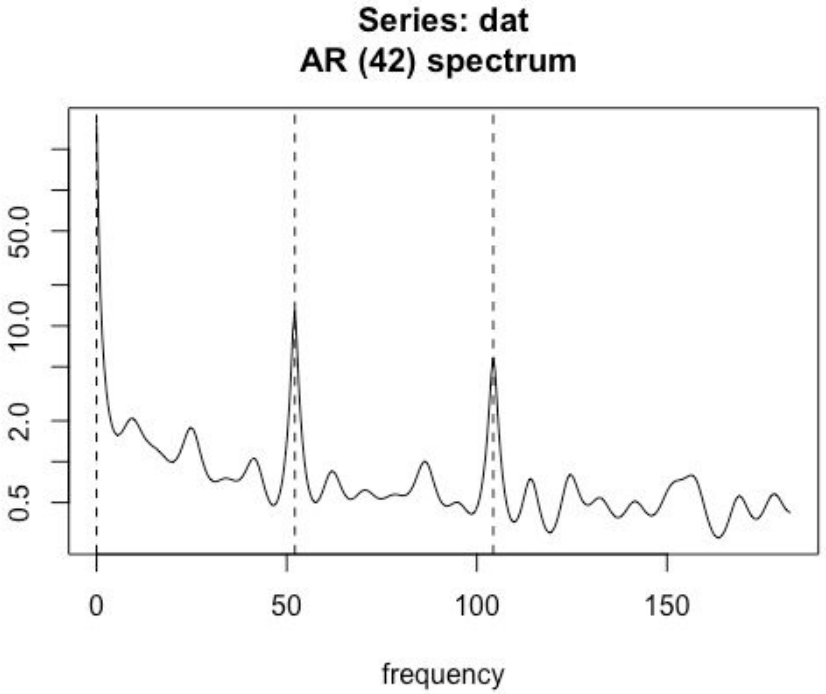
Spectral Analysis





Spectral Analysis





Next Steps

Forecasting with long seasonal periods

It seems that the data has 365 long cycles from the seasonal decomposition.

Choose Fourier series approach to forecast the data.

Look at other potential covariates

Traffic volume

Add area information