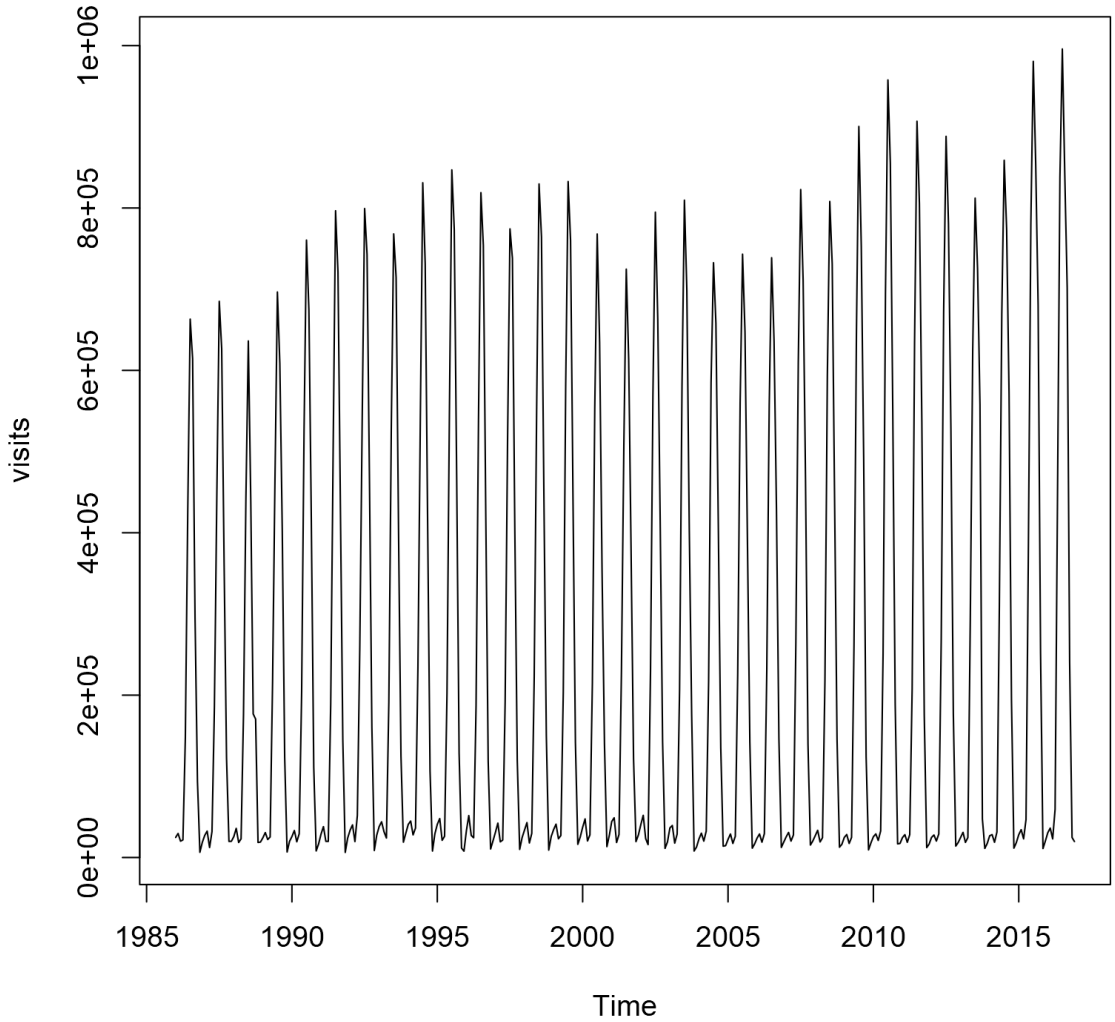
Time Series Analysis Project: Yellowstone Visitor Trends

Michael Dunn, Jacob Klein, and Evan Waldmann

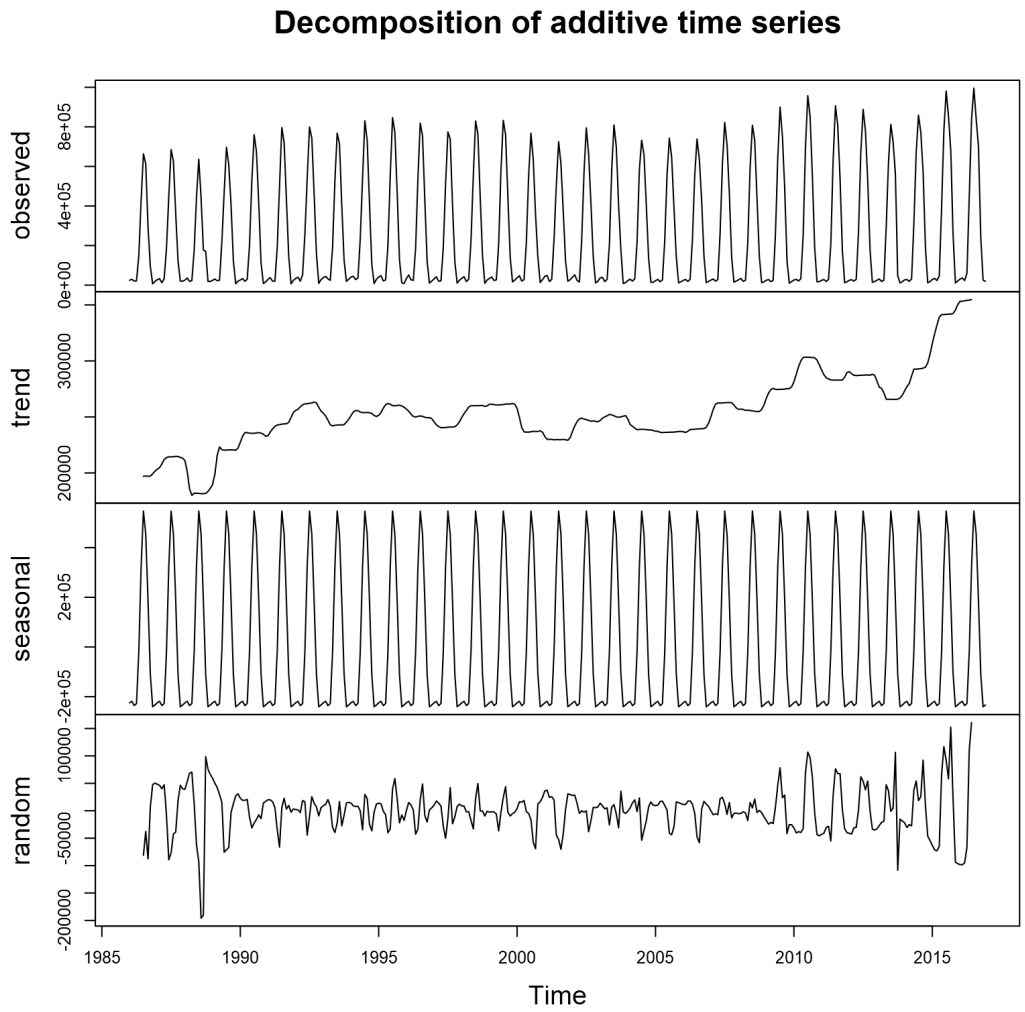
University of Central Florida

April 18th, 2019

Yellowstone visitor attendance 1986-2017

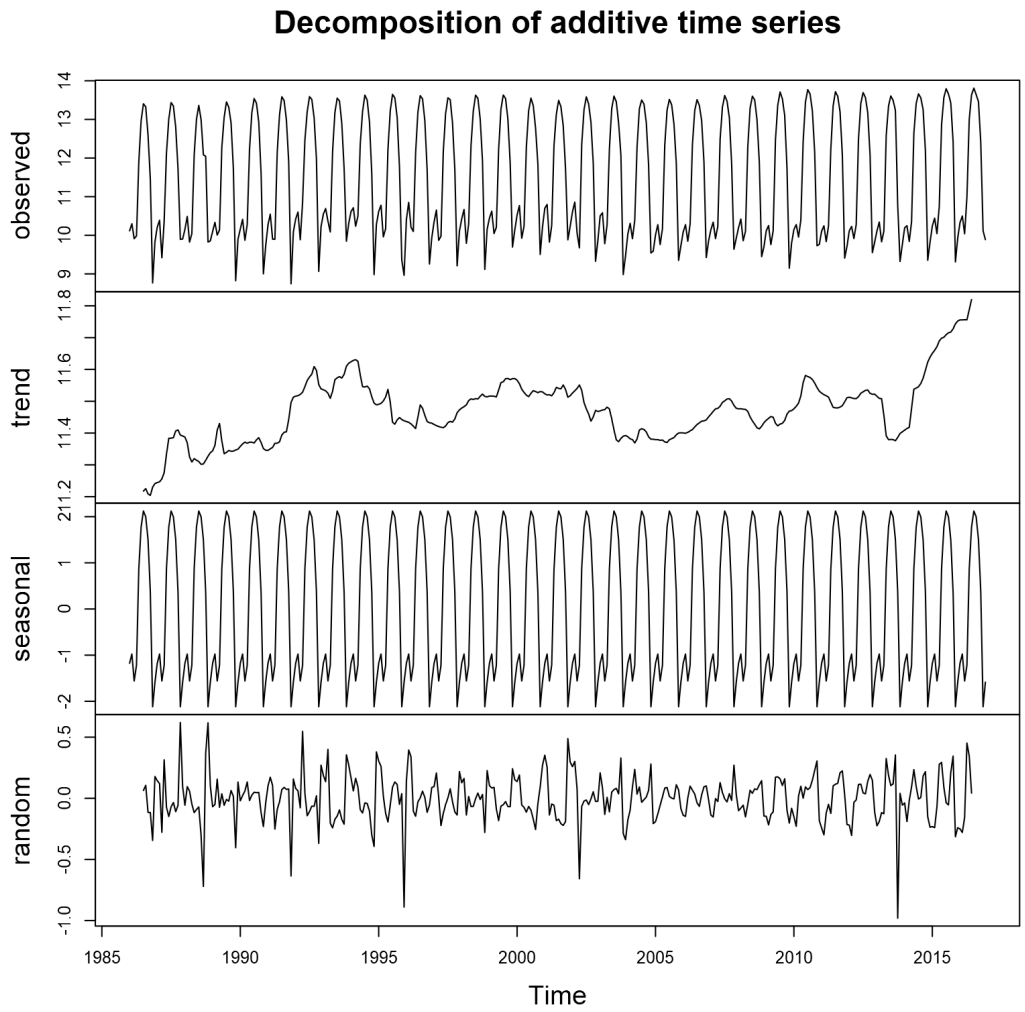


General trends of the time series



Note the linear trend, heavy seasonality, and 1988 outlier

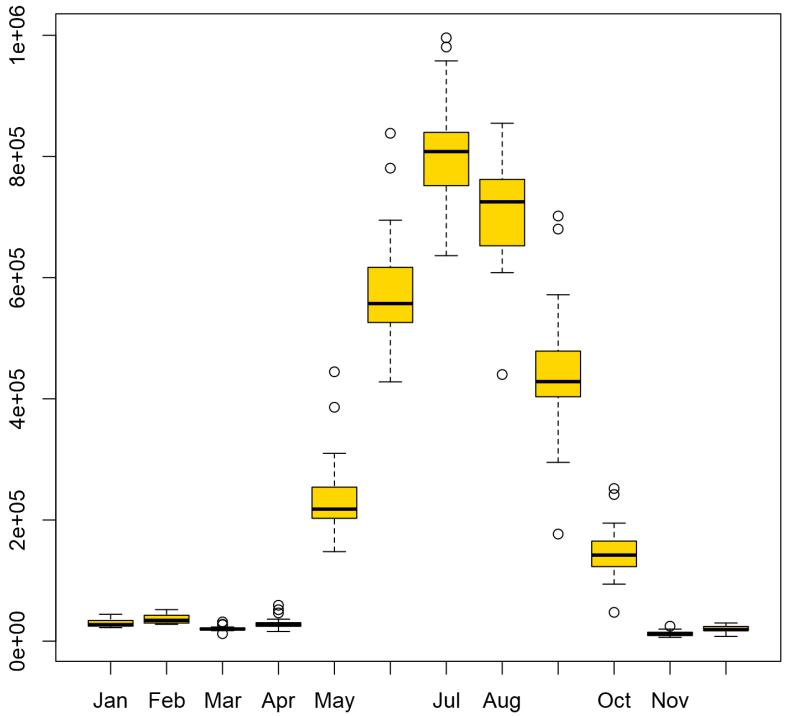
Log transform to reduce linear trend



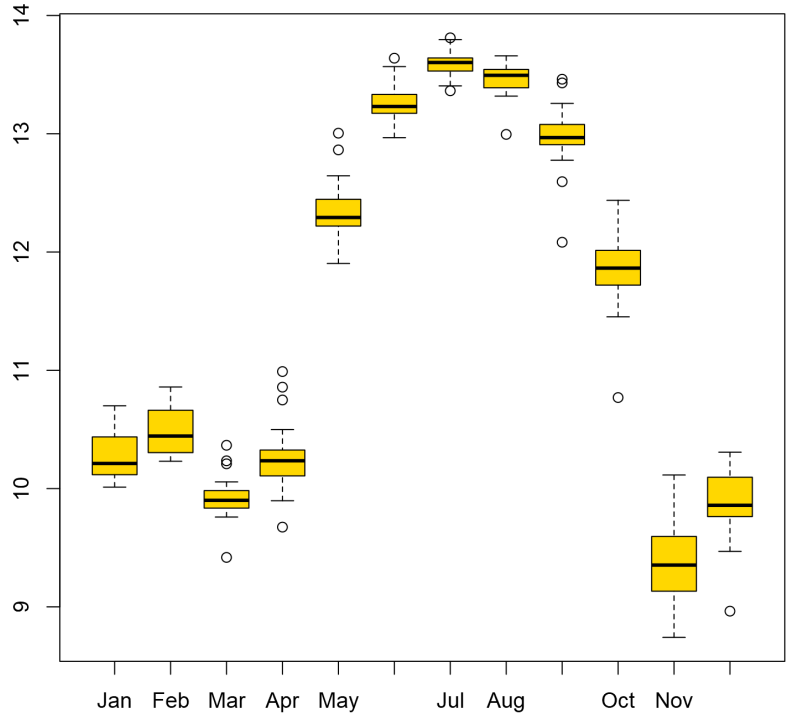
With the transform the trend is decreased, but we have more peaks in our random error

Seasonality

Distribution without transformation



Distribution with log transformation

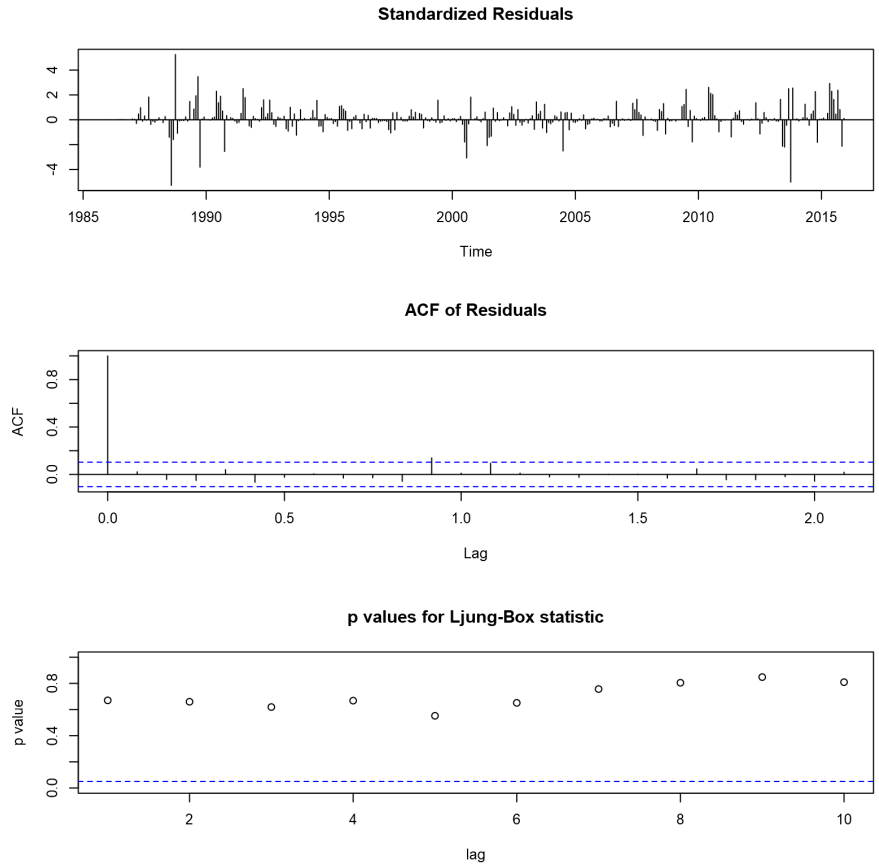


Summer months have drastic increase in visitors

Model Choice

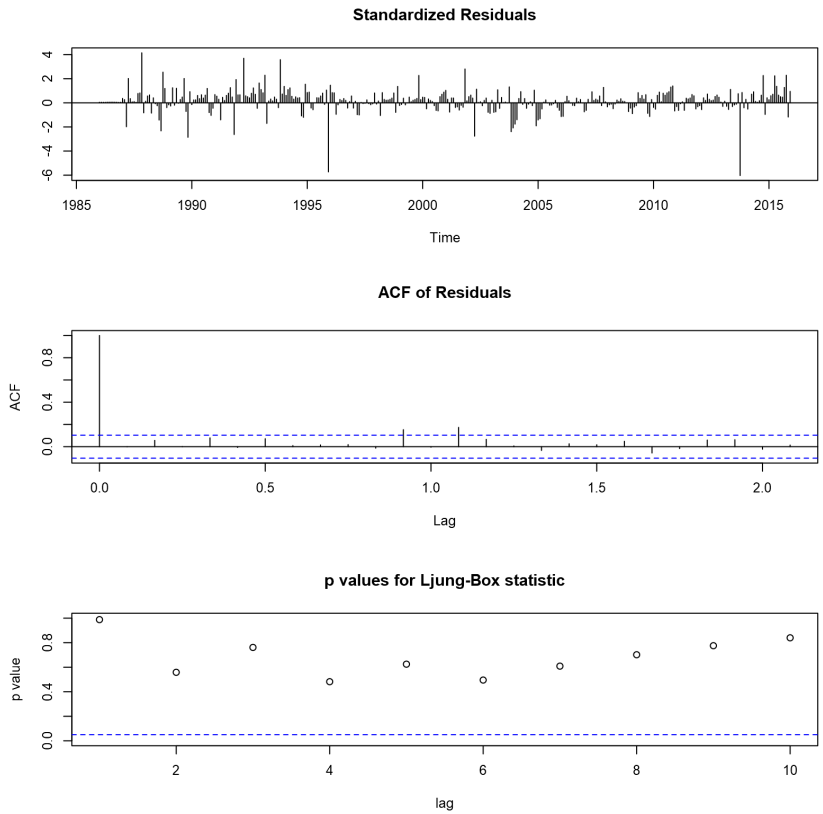
- Both series had p values of less than .01 with Augmented Dickey-Fuller Test
- Held last year of data from each series in order to check the accuracy of our forecasts
- Systematically chose models with the lowest AIC value for our two series to get the following
- **Regular data:** AR(1) model with a seasonal component of order (0,1,1)
- **Log transform:** MA(1) model with a seasonal component of order (2,1,1)

Model Diagnostics: ARIMA (1,0,0) X (0,1,1) [data without transformation]



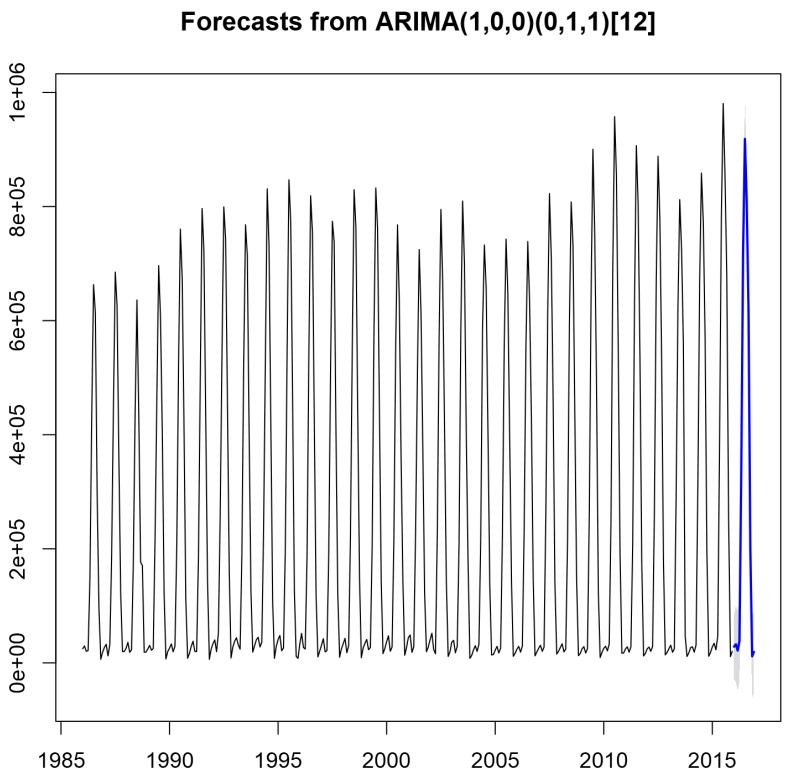
Again note the 1988 outlier and that the Ljung-Box statistic is significant up to a lag of 15

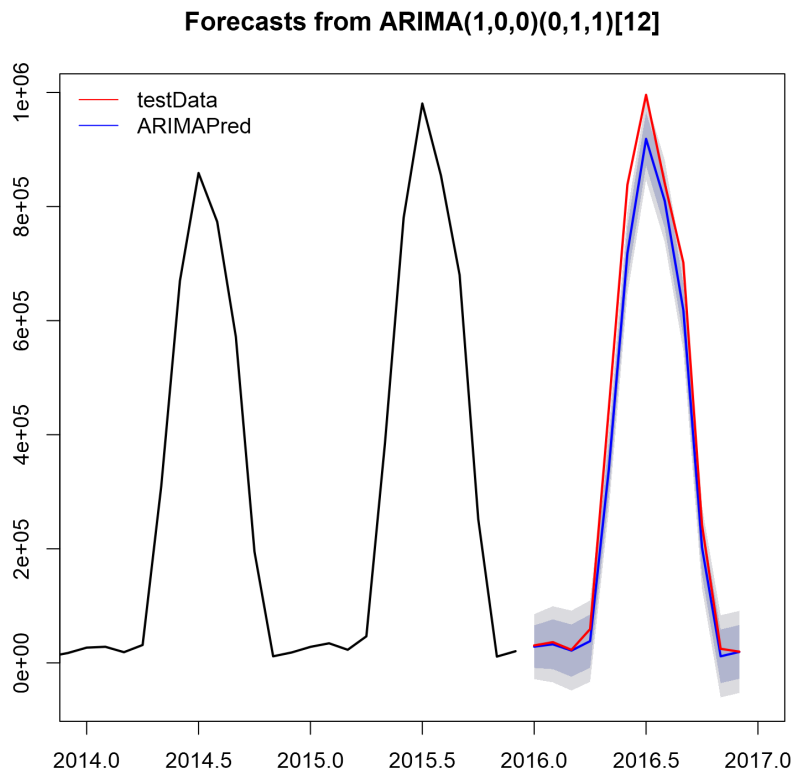
Model Diagnostics: ARIMA (0,0,1) X (2,1,1) [data with transformation]



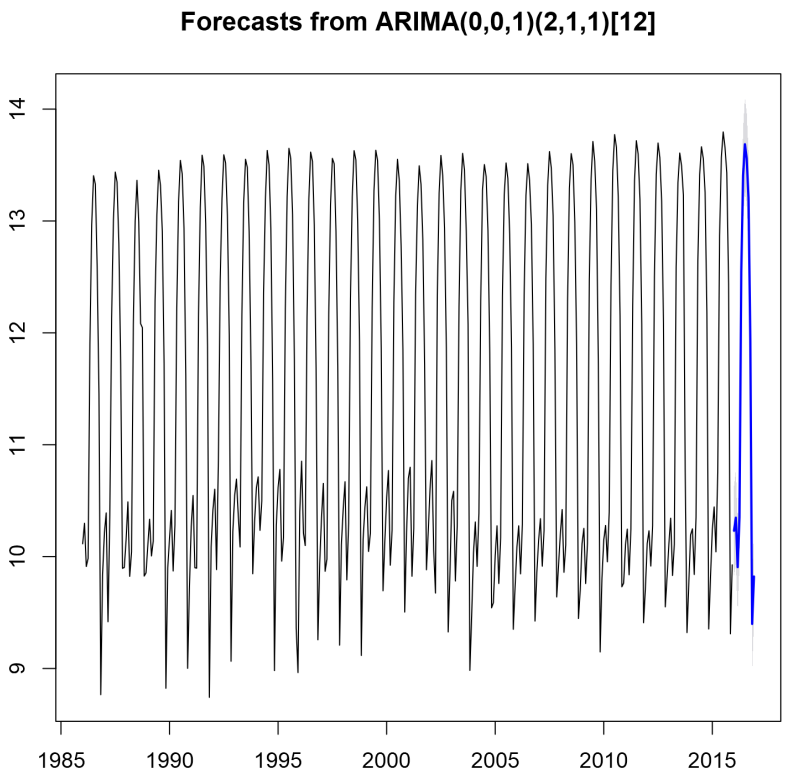
With the transform the trend is decreased, but we have more peaks in our random error. Also the Ljung-Box statistic for the transformed data is only significant up to a lag of 12

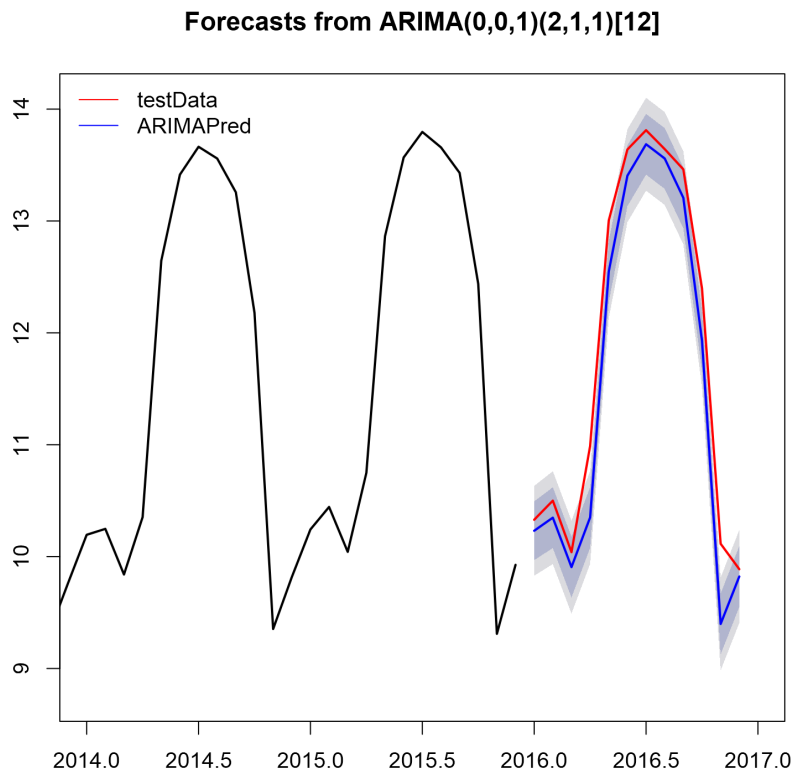
Forecast of model without log transformation





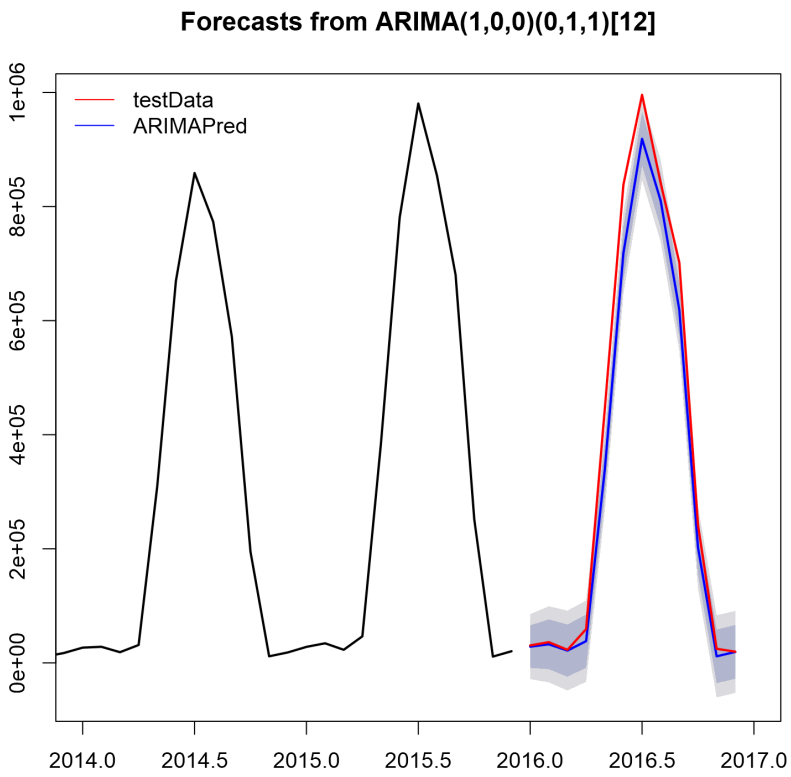
Forecast of model with log transformation



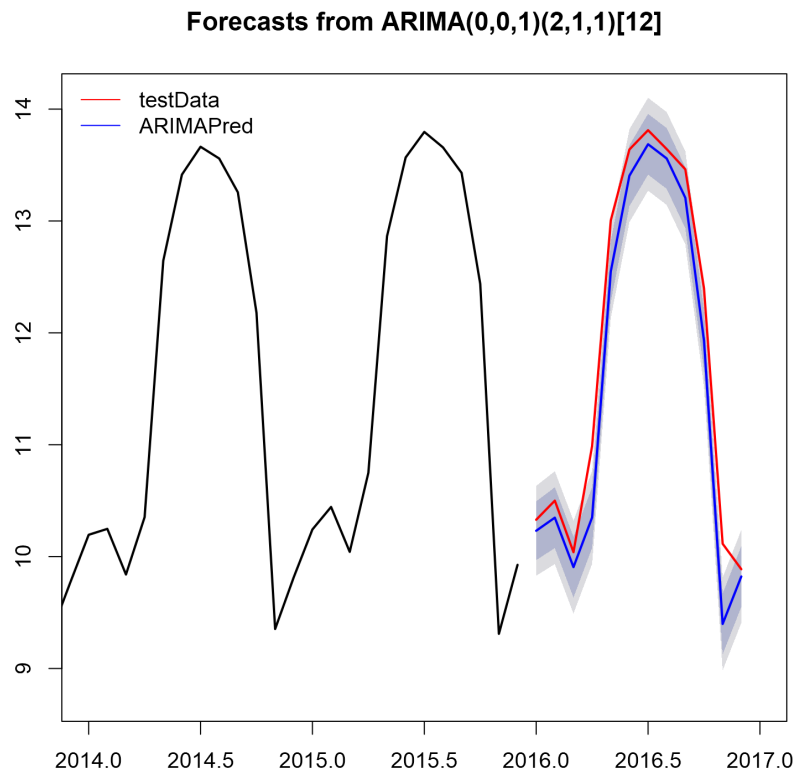


Comparison of the two forecasts

Model on regular data



Model on log transformation



We again see that the non-log transformed model has a lower U value of 0.2662 than the log transformed model's U value of 0.3558.

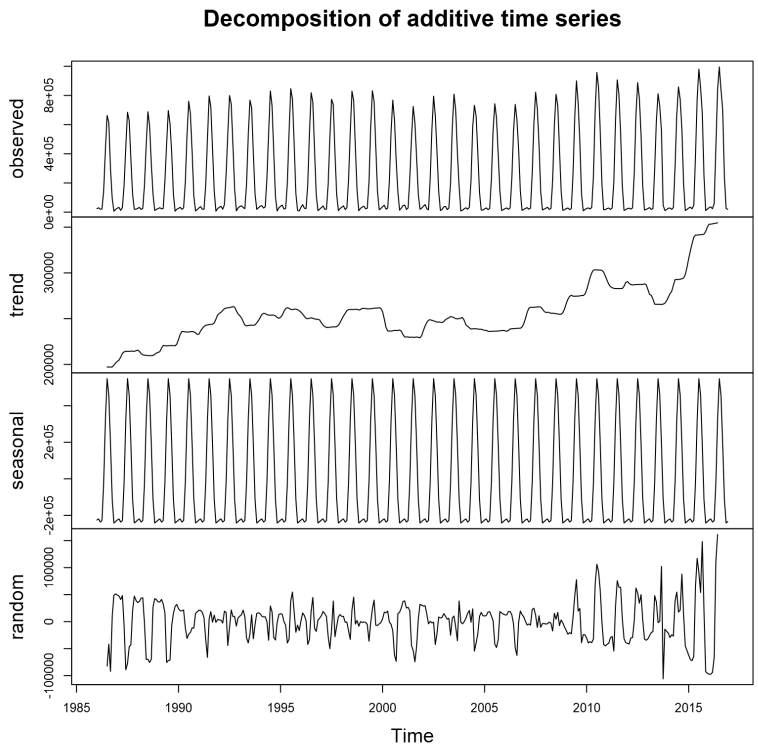
1988 Wildfire

Yellowstone fires of 1988 collectively formed the largest wildfire in the recorded history of Yellowstone National Park in the United States.

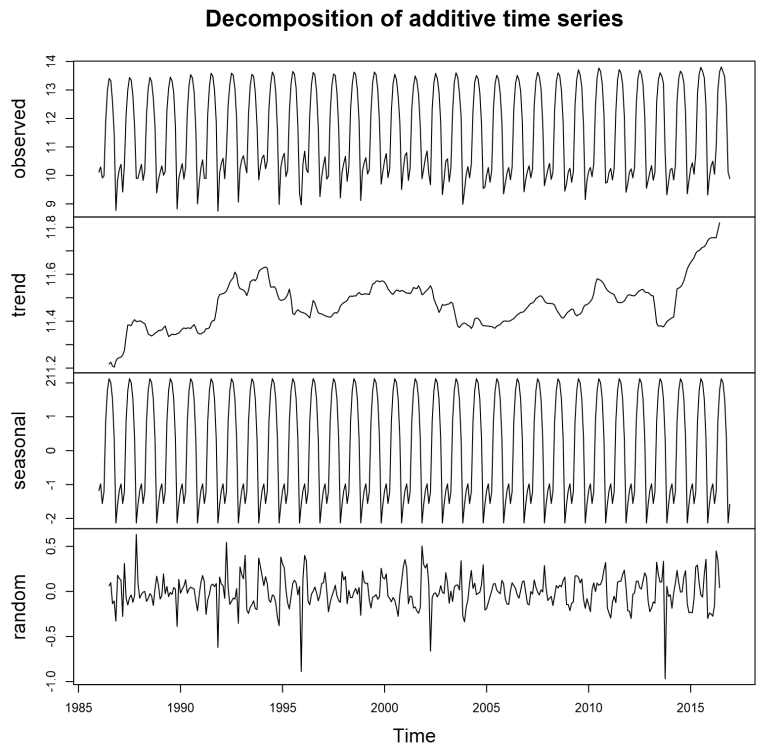
Fires approach the Old Faithful Complex on September 7, 1988

Decomposition after replacing 1988 with 5 year average

Regular data



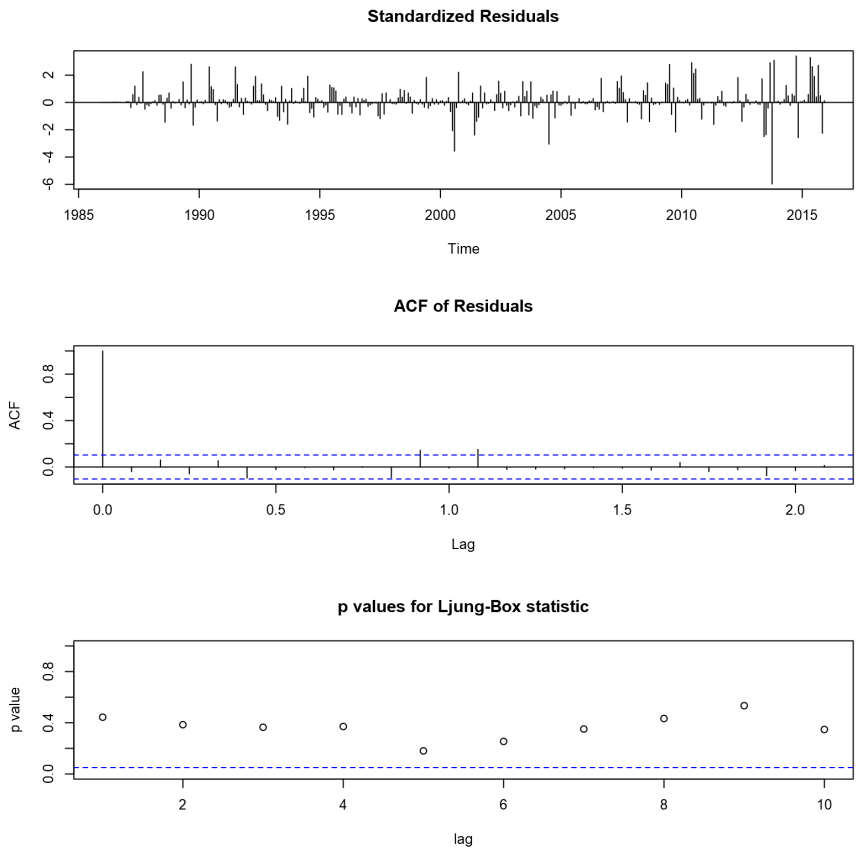
Log transformation



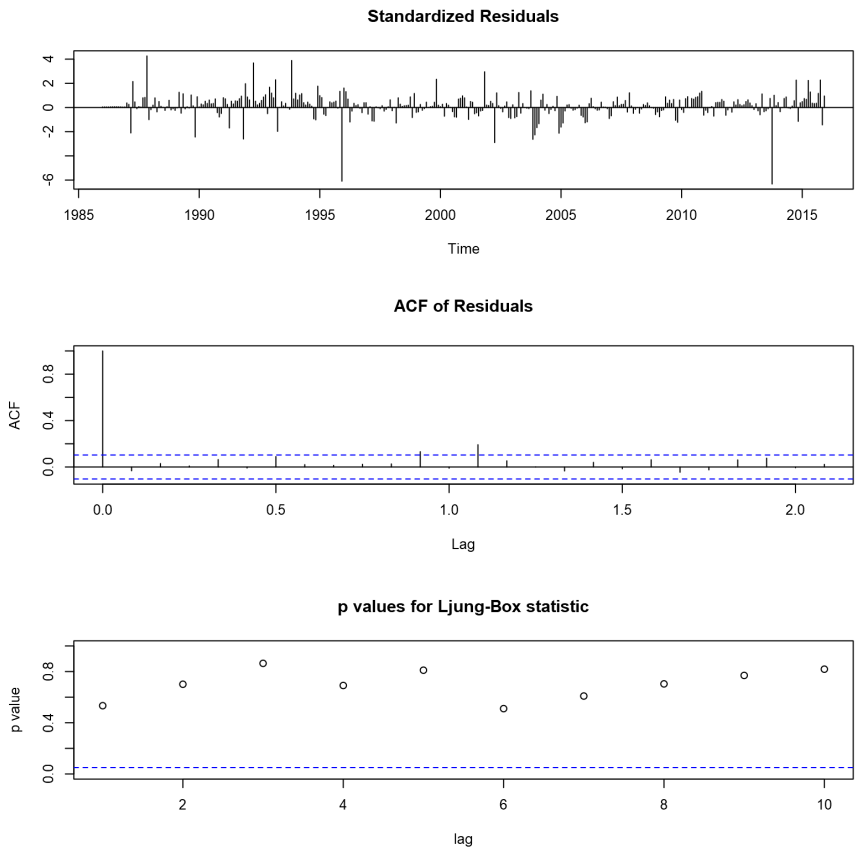
The data manipulation fixed the outlier, and the log transforms still lessens the linear trend and decreases variance of the random errors.

New model fitting and diagnostics

Model on regular data



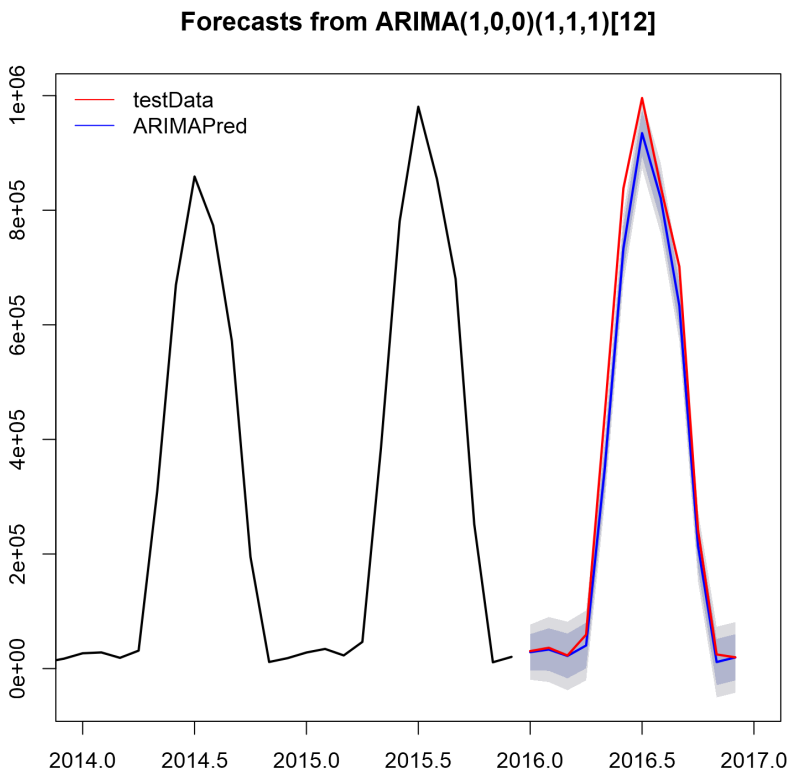
Model on log transformation



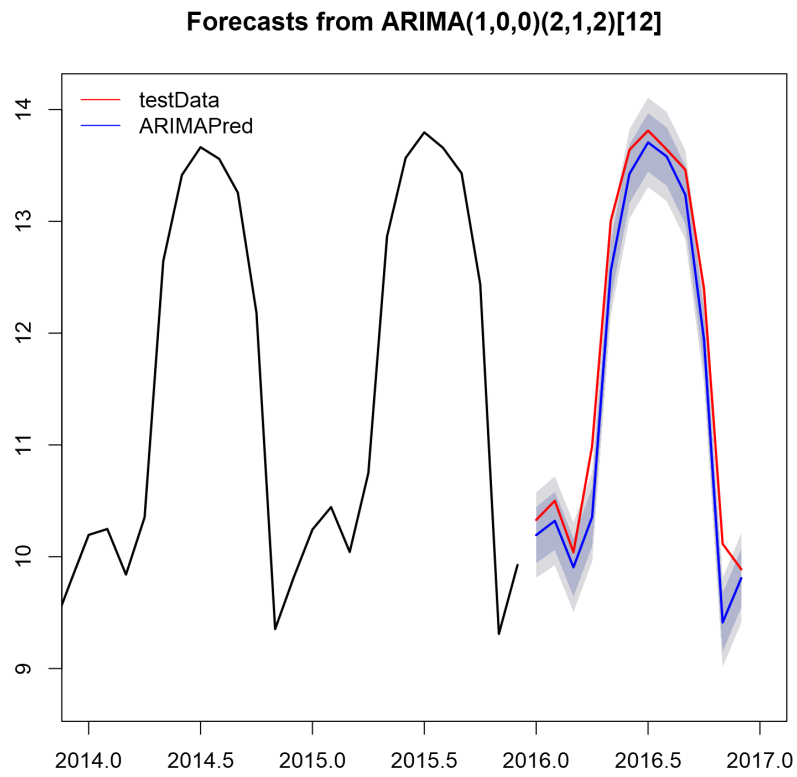
The original data only has a larger outlier at 2014 now, but the log transformed data as a couple outliers.

Forecasting with the new models

Model on regular data



Model on log transformation



We again see that the non-log transformed model has a lower U value of 0.2662 than the log transformed model's U value of 0.3558.

Conclusion

- We first presented two stationary time models and found that found that ARIMA (1,0,0) X (0,1,1) model was best suited for forecasting.
- We provided justification of the used of a seasonal model as well as explore some model diagnostics for what we presented.
- After manipulating our data to account for our largest outlier, we concluded the best model for forecasting is the ARIMA(1,0,0) X (1,1,1) with the modification of the fire year average to replace the 1988 visitor attendance because of the large wildfire.