May 2018

Modeling and Forecasting Crime Patterns in Bellingham, Washington

Zachary Domingo
Western Washington University

Eric Shoner
Western Washington University

Modeling and Forecasting Crime Patterns in Bellingham, WA

Zak Domingo, Eric Shoner, Phillip Thomas, Justin Irvine, advised by Kimihiro Noguchi

Department of Mathematics, Western Washington University, Bellingham, WA 98225

WESTERN WASHINGTON UNIVERSITY

Abstract:

Our purpose is to use time series analysis to model and forecast the underlying dynamics behind crime in Bellingham, Washington. Using recent monthly data from the Bellingham Police Department, we considered singular spectrum analysis (SSA) and autoregressive integrated moving average (ARIMA) modeling techniques to estimate significant deterministic patterns in the data. We created two models for alcohol offenses and domestic violence datasets then compared them. The better performing model was used to forecast the number of crime incidents for the next ten months and identify trends and seasonality.

Background and Significance:

The federal and state governments spend billions of dollars on maintaining law enforcement agencies. If we had the ability to predict trends or periodicities in crime patterns, these agencies could then properly allocate their resources to reduce the incidence. This would allow the local law enforcement to take more meaningful action or make better informed decisions in measures to reduce crime. This research takes a somewhat narrow scope, considering monthly crime data from Bellingham, Washington (Fig, 1, Fig. 2) [1].

Research Questions:

- How well can we model crime trends/seasonalities in Bellingham, specifically, alcohol offenses and domestic violence?
- Which of two models (SSA-based or ARIMA) performs better in terms of forecasting?

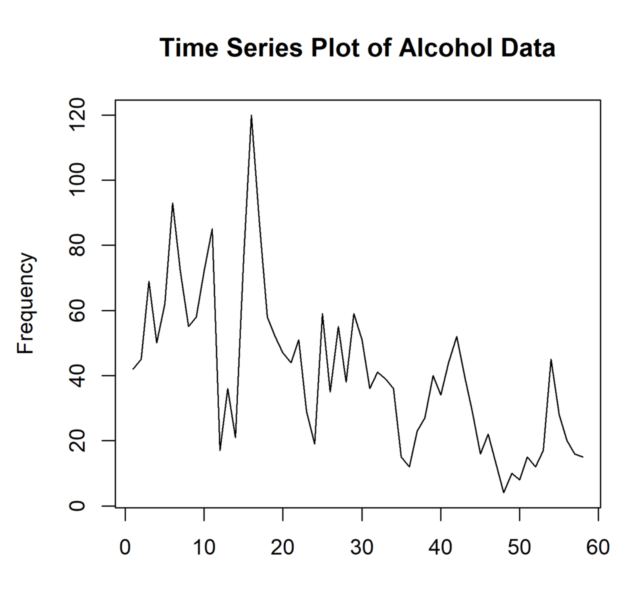

Figure 1. Time series plot of alcohol crime data.

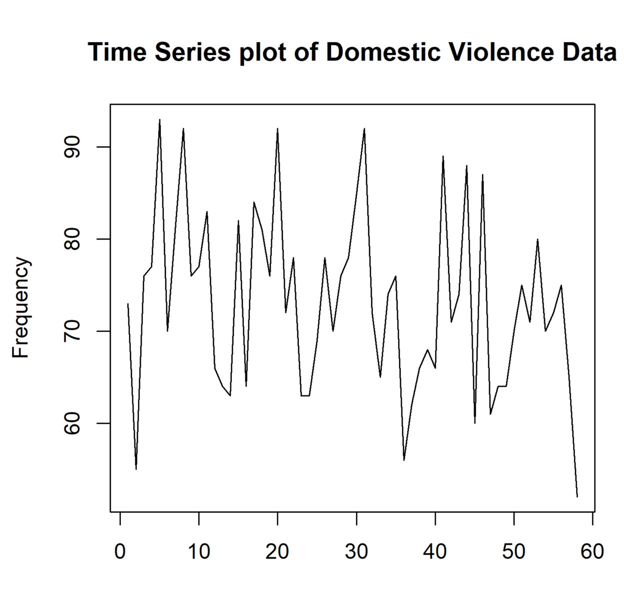

Figure 2. Time series plot of domestic violence data.

Summary of Statistical Models:

- SSA is a relatively new method that can model the deterministic component (trend and seasonality) simultaneously without any restrictive assumptions [2]. If the residuals of the SSA are correlated, then they may be modeled by the traditional autoregressive moving average (ARMA) model.
- ARIMA is a typical time series model that handles non-stationary time series data by differencing [3].

Methods:

- Compared an ARIMA model with a combination of SSA and ARMA model (ARIMA without any differencing) for each data set.
- Performed a logarithmic transformation on all data sets to ensure appropriate comparisons.
- Removed one outlier in alcohol data set.
- Used the first 48 observations (first 4 years) to create our models so that we could forecast the next 10 (9 if an outlier was removed) months and compared them to observations from months 49 to 58 (57 with an outlier removed).
- Used the Diebold-Mariano test (D-M test) and root mean square error (RMSE) values to see if one model performed better than the other per dataset.

ARIMA Modeling and Forecasting for Domestic Violence:

- Using the first 48 observations as training data, the AICc criterion selected an ARIMA(2,1,0) model [3].
- Forecasted values using the ARIMA(2,1,0) model for the trained data with 80% and 95% prediction intervals (Fig. 3).
- Examined the autocorrelation function (ACF) and partial ACF (PACF) plot, and applied a Portmanteau test on the residuals of the model to ensure the residuals were uncorrelated (Fig. 4).
- Prediction intervals from model capture the data.

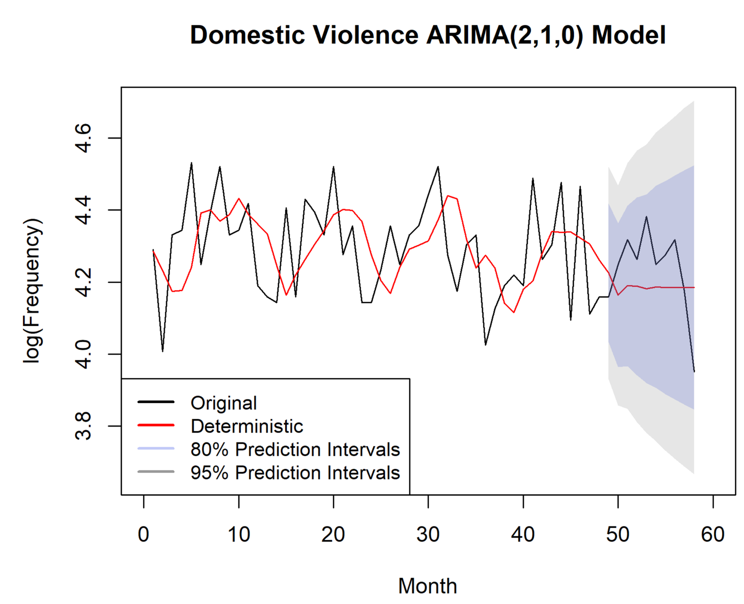

Figure 3. Domestic violence data with ARIMA(2,1,0) forecasts and prediction intervals.

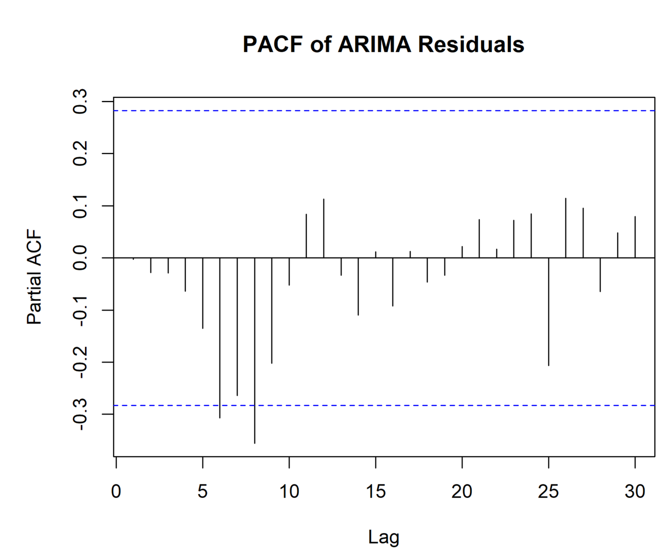

Figure 4. Domestic violence ARIMA(2,1,0) model PACF plot of residuals.

ARIMA Modeling and Forecasting for Alcohol Offenses:

- Using the first 47 observations as training data, the AICc criterion selected a first-order autoregressive (AR (1)) model [3].
- Forecasted values using the AR(1) model for the trained data with 80% and 95% prediction intervals (Fig. 5).
- Examined the autocorrelation function (ACF) and partial ACF (PACF) plot, and applied a Portmanteau test on the residuals of the model to ensure the residuals were uncorrelated (Fig. 6).
- Model prediction intervals capture the observed data set.

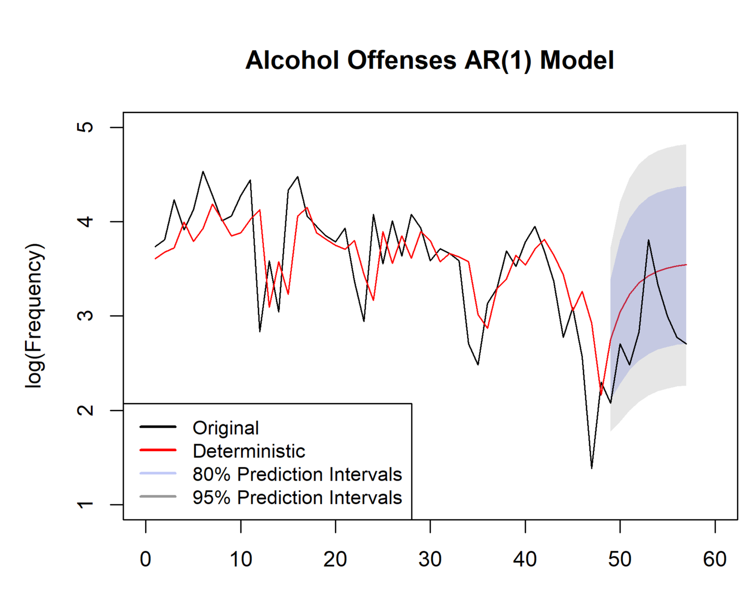

Figure 5. Alcohol crime data with AR(1) forecasts and prediction intervals.

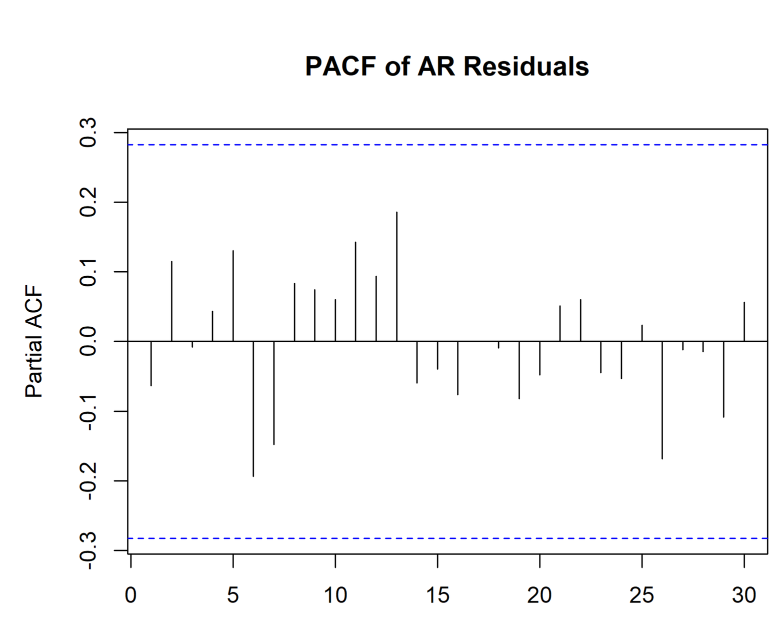

Figure 6. Alcohol crime AR(1) model PACF plot of residuals.

SSA-Based Modeling and Forecasting for Domestic Violence:

- The seasonality patterns were estimated to have around 12- and 2.5-month periods, implying a yearly component and a rough quarterly periodicity following the change of the seasons.
- Fitted an ARMA(1,1) model on the residuals of the SSA model based on the ACF and PACF plots.
- Ensured that the residuals after SSA + ARMA were more or less uncorrelated by the Portmanteau test, ACF, and PACF plots.
- Forecasted values using the SSA + ARMA model for the trained data with 80% and 95% prediction intervals (Fig. 7).
- All but one of the points fell within the 80% prediction interval (Fig. 7). The one outlying point seems to follow the general downward trend of the data, causing no immediate concern pertaining to forecast performance.

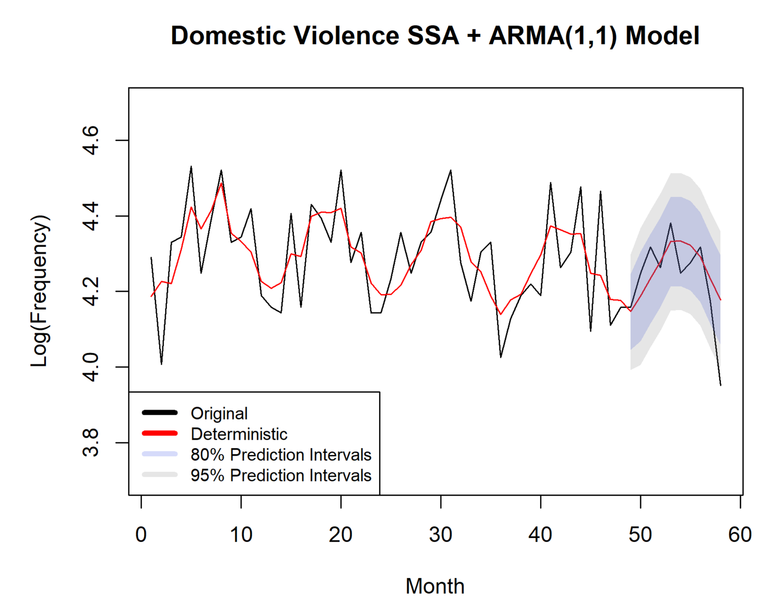

Figure 7. Domestic violence data with SSA + ARMA(1,1) forecasts and prediction intervals.

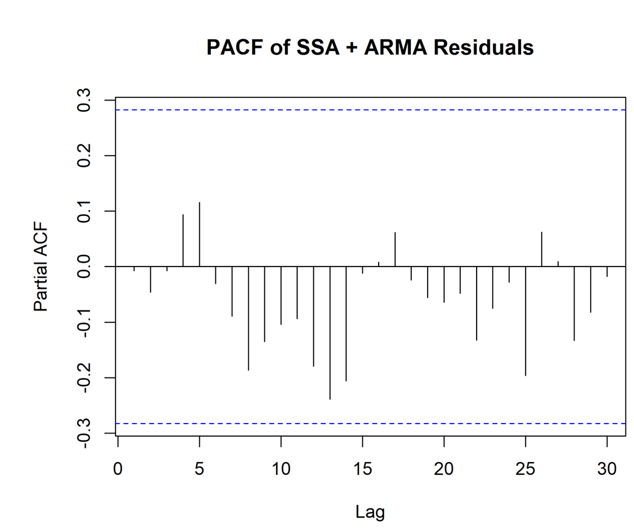

Figure 8. Domestic violence data with SSA + ARMA(1,1) model PACF plot of residuals.

SSA-Based Modeling and Forecasting for Alcohol Offenses:

- The seasonality patterns were estimated to have around 12- and 4-month periods, implying a yearly component and roughly the amount of time between holidays associated with increased alcohol consumption, i.e., Fourth of July, Halloween, etc.
- Examined that the residuals of the SSA model seem uncorrelated by checking the autocorrelation function (ACF) and partial ACF (PACF) plot. Thus, no ARMA model was fitted for the residuals after SSA.
- Most of observed data not captured in prediction interval, but appears to follow forecast trend.

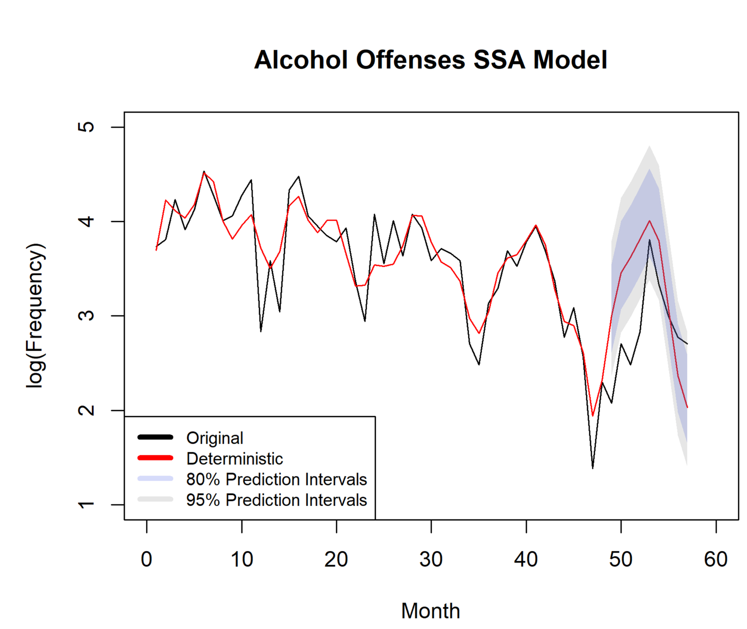

Figure 9. Alcohol crime data with SSA model forecasts and prediction intervals.

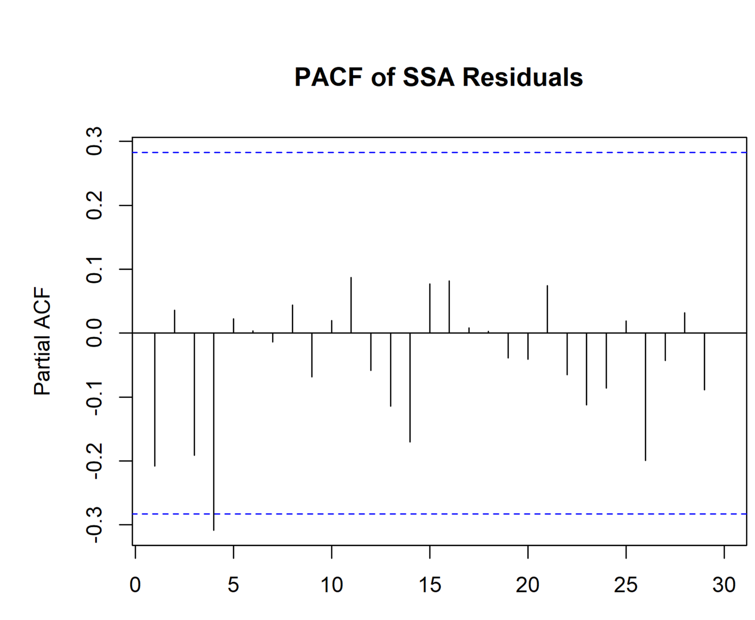

Figure 10. Alcohol SSA model PACF plot of residuals.

Analysis on Forecasting Performance:

Domestic Violence:

- The RMSE values for the SSA model are lower, suggesting its superior forecast performance (Fig. 11)
- On the other hand, the D-M test shows no significant difference in performance be tween the two models for all forecast values except one, implying a weak advantage of the SSA model (Fig. 12).

Alcohol Offenses:

- The RMSE values for the SSA model are lower, suggesting its superior forecast performance (Fig. 13).
- On the other hand, the D-M test does not indicate any statistically significant advantage of the SSA compared to the AR(1) model (Fig. 14).

Overall:

- Based on the RMSE values alone, the SSA-based models seem to predict the observed values with greater accuracy compared to the ARIMA models.
- The size of the data set may have influenced performance of all models. Specifically, the SSA-based models performed well with small data sets whereas the ARIMA models would require a larger data set.

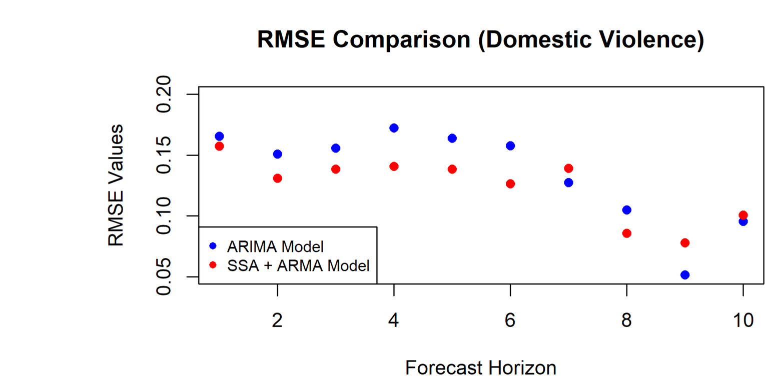

Figure 11. RMSE values of forecasts from domestic violence models.

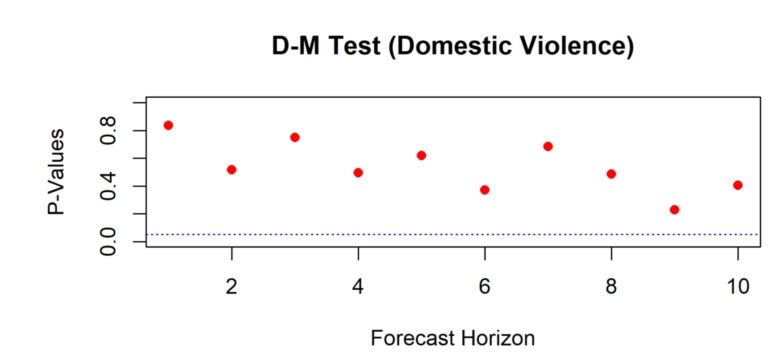

Figure 12. P-values from DM test of domestic violence models.

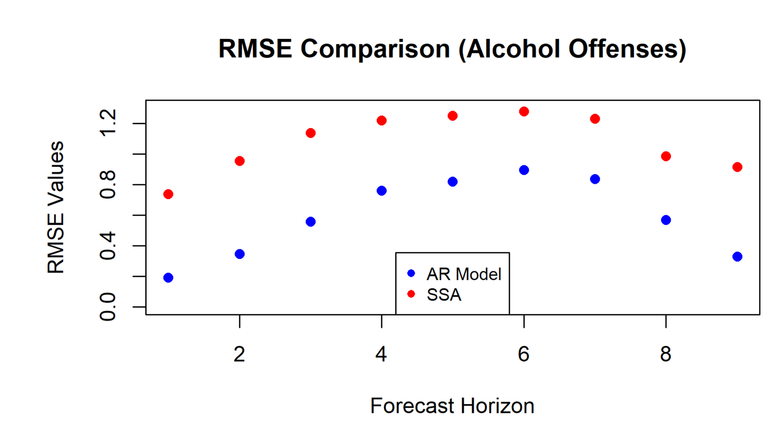

Figure 13. RMSE values of forecasts from alcohol crime models.

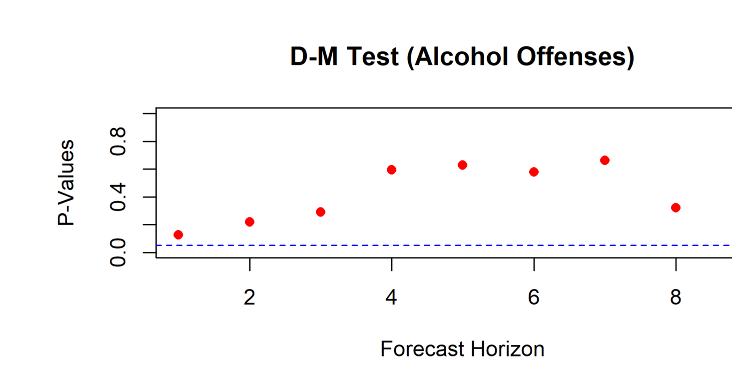

Figure 14. P-values from DM test of forecasts from alcohol crime models.

Conclusions:

The SSA-based models tend to produce much more accurate forecasts than the ARIMA models for the two data sets being analyzed. These results may be attributed to the fact that the SSA can be used even for relatively small data sets.

The SSA-based models for both data sets indicate yearly periodicity for both datasets. Both data sets show increases in frequency during winter and peaks in summer months, and decreases for the rest of the year.

The above finding could be attributed to the fact that children are not in school and thus are more likely to become the victim of domestic violence. Similarly, the increase in alcohol crimes could be attributed to minors seeking out alcohol when school is not in session. Further investigation is required to substantiate these claims and it is likely that unaccounted for factors contribute to these periodicities.

To further this research, a multivariate analysis of correlation between domestic violence crimes and alcohol crimes over time may be considered.

References:

[1] Bellingham Police Department (2017). Crime Statistics [Data table for each year 2013 -2017]. Retrieved from https://www.cob.org/gov/dept/police/news/Pages/crime-stats.aspx

[2] Golyandina, N. & Korobeynikov, A. (2014). Basic singular spectrum analysis and forecasting with R. *Computational Statistics and Data Analysis*, 71, 934-954.

[3] Shumway, R. H. & Stoffer, D. S. (2016). *Time Series Analysis and Its Applications with R Examples*, 4th Ed. Springer.