

Santa Clara Crime Data Analysis

- Stats 626 Time Series TAMU Summer 2024
- Team 4

Outline

- Background and explanation of data
- Data exploration and analysis of trends
- Models and diagnostics for monthly & weekly counts
- Models and diagnostics for daily counts
- Models for hourly counts
- Conclusions and recommendations

Background on Santa Clara County

- Overview
 - Known as "Silicon Valley"
 - Located in Northern California
 - Home to tech giants like Apple, Google, and Facebook
 - County seat: San Jose
- Geography
 - Area: ~1,304 square miles
 - Bordered by San Mateo, Alameda, Stanislaus, Merced, and Santa Cruz counties
 - Includes the Santa Clara Valley
- Demographics
 - Population: ~2 million
 - Highly diverse and educated population
 - High cost of living
- Economy
 - Driven by the tech industry
 - Significant presence of software, hardware, and biotech companies
 - High average incomes, but challenges with housing affordability and income inequality

Recent News

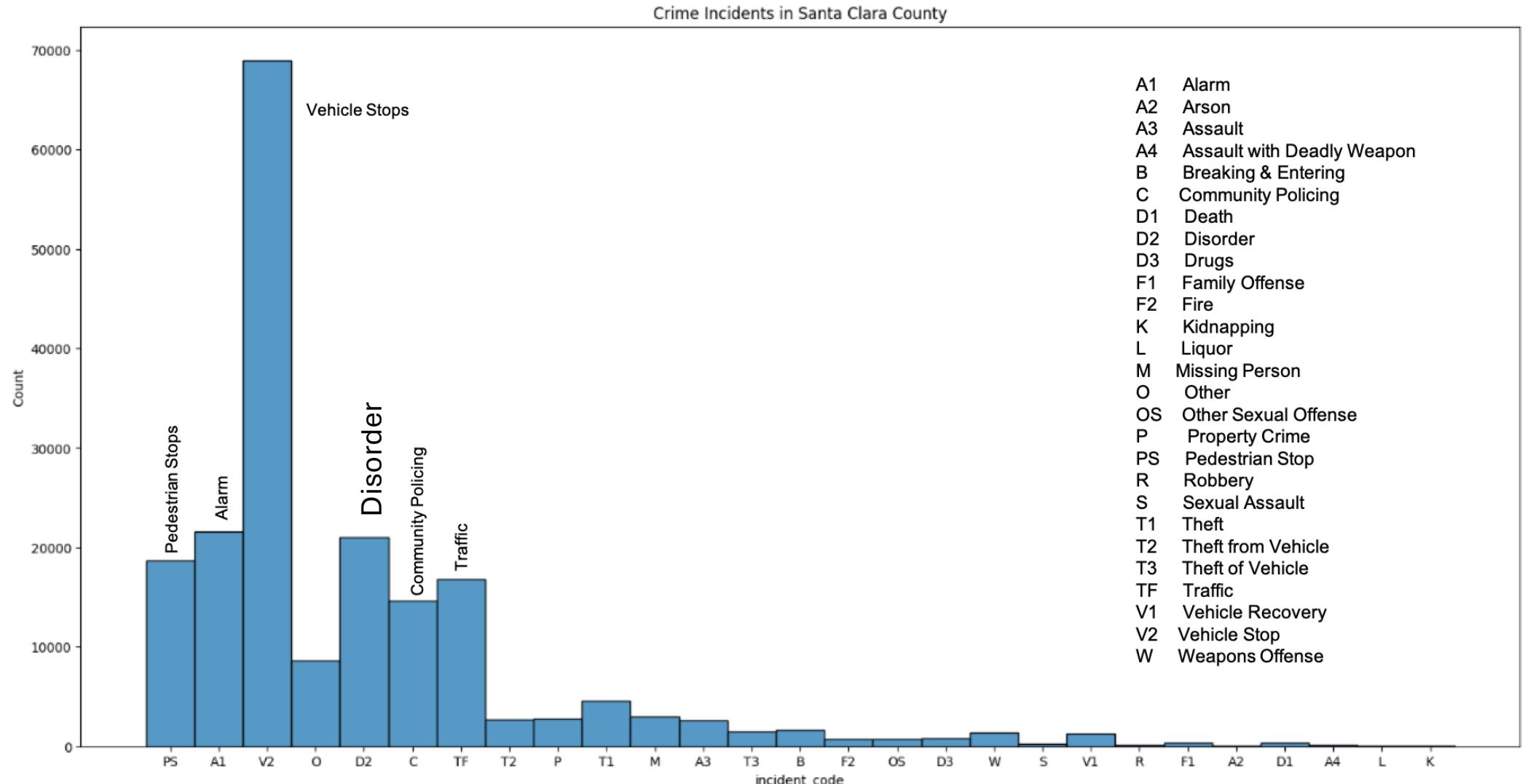
- Overwhelming caseloads since COVID
- Issues with scandals and corruption in the past few years
- Attempts at reforming crime prevention efforts (in progress)
 - New monitoring devices in addition to ankle monitors
 - Installing new speeding cameras across San Jose

While our dataset spans a time period prior to these events, it helps provide context and shed light on what was happening just before Covid

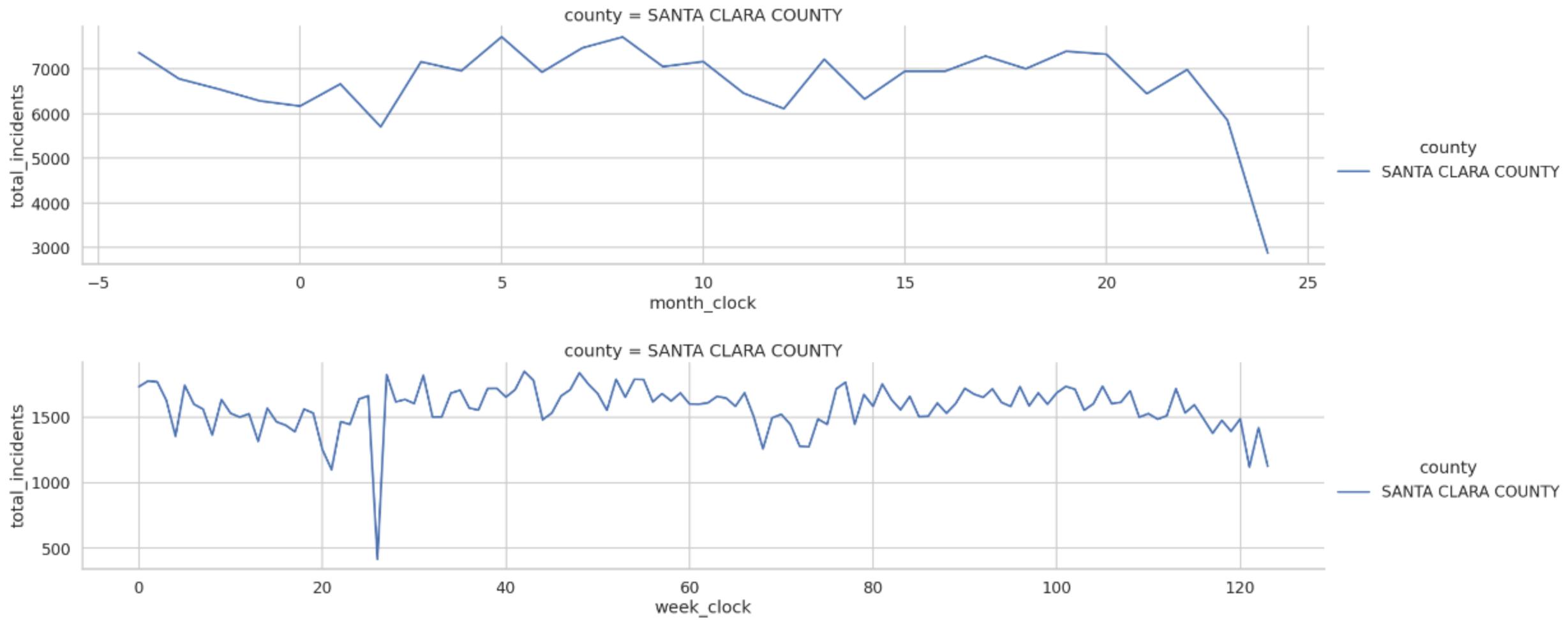
Santa Clara County Sheriff's Office Incident Dataset

- Dataset Overview
 - Contains crime reports from Aug 2017 to Dec 2019
 - Covers all reported incidents to the Sheriff's Office
 - **Date and Time:** Specific dates and times of incidents
 - **Incident Types:** Categories such as theft, assault, vandalism, etc.
 - **Locations:** Geographic locations of crimes
 - **Other Details:** Additional relevant information
- Analysis Focus
 - on preventive policing incidents, each with more than 10 counts daily
 - Vehicle Stops - 68,912 ~ 35.4%
 - Alarms - 21, 563 ~ 11.1%
 - Disorder - 20,981 ~ 10.8%
 - Pedestrian Stops - 18,652 ~ 9.6%
 - Traffic - 16,768 ~ 8.5%
 - Community Policing – 14,635 ~7.5%

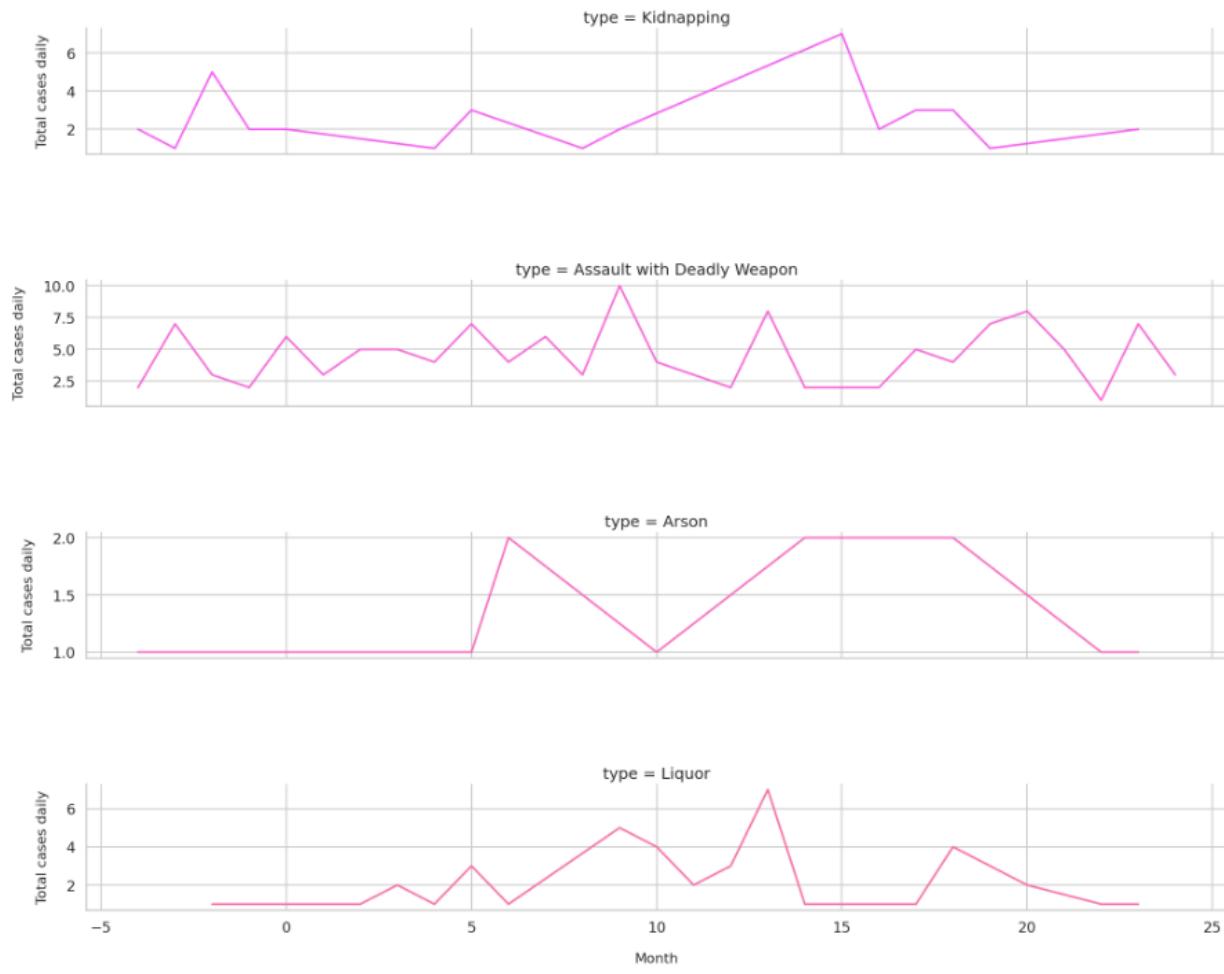
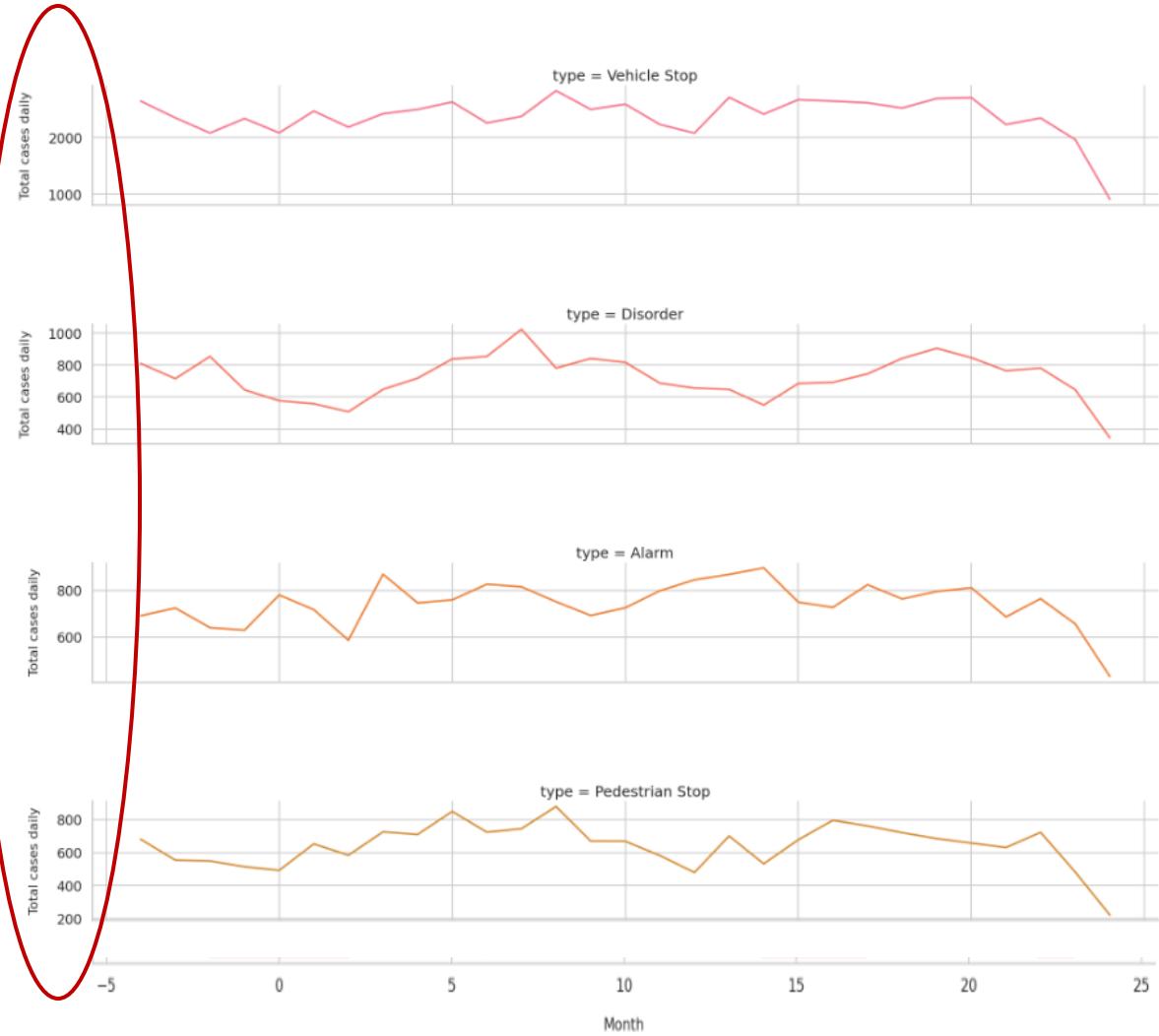
Santa Clara County Sheriff's Office Incident Dataset



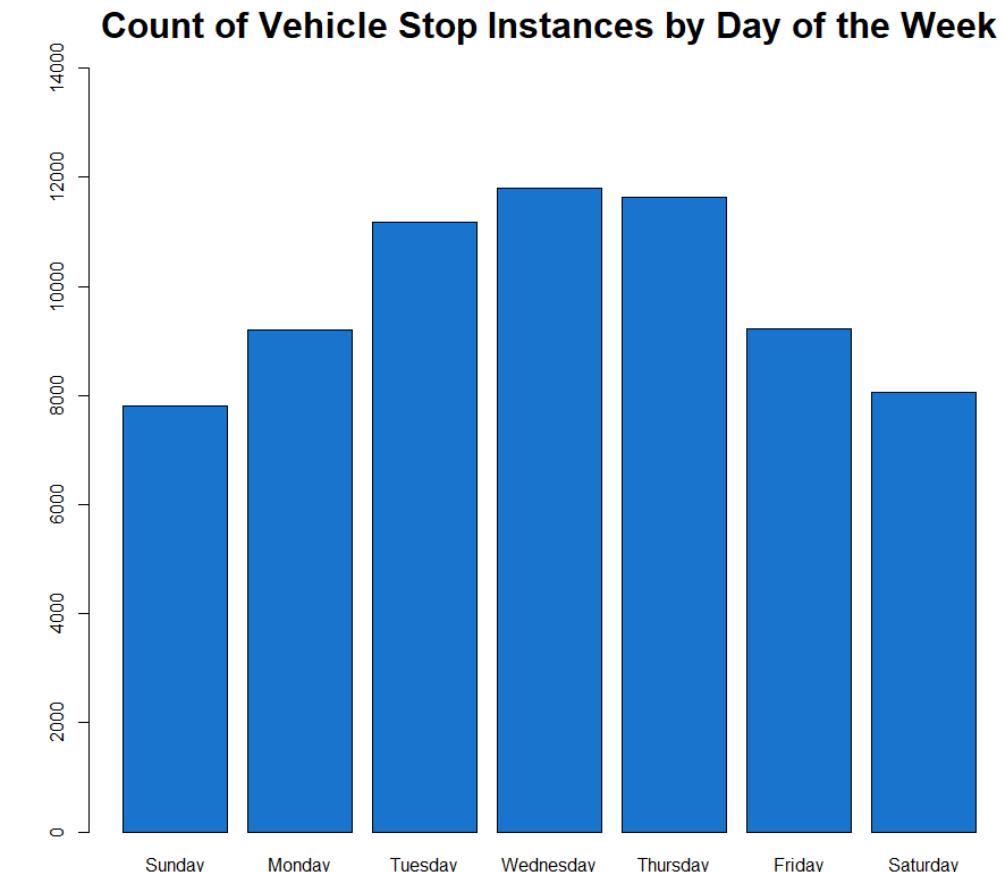
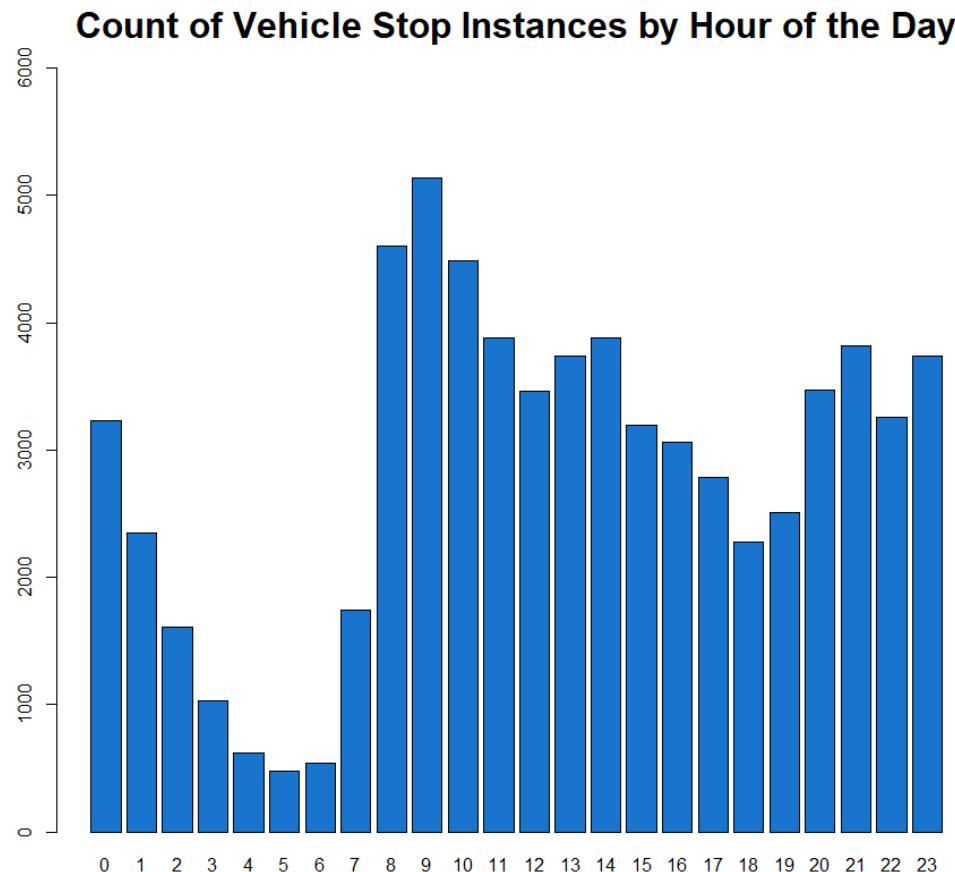
Monthly & Weekly Trends



Monthly Trends

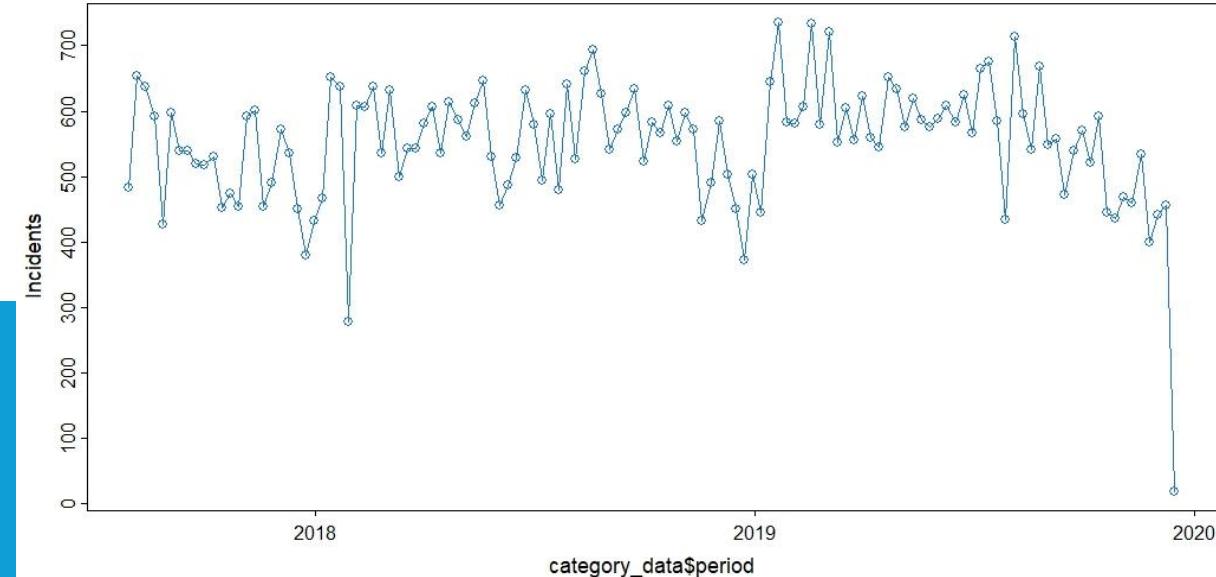


Vehicle Stop Instance Trends

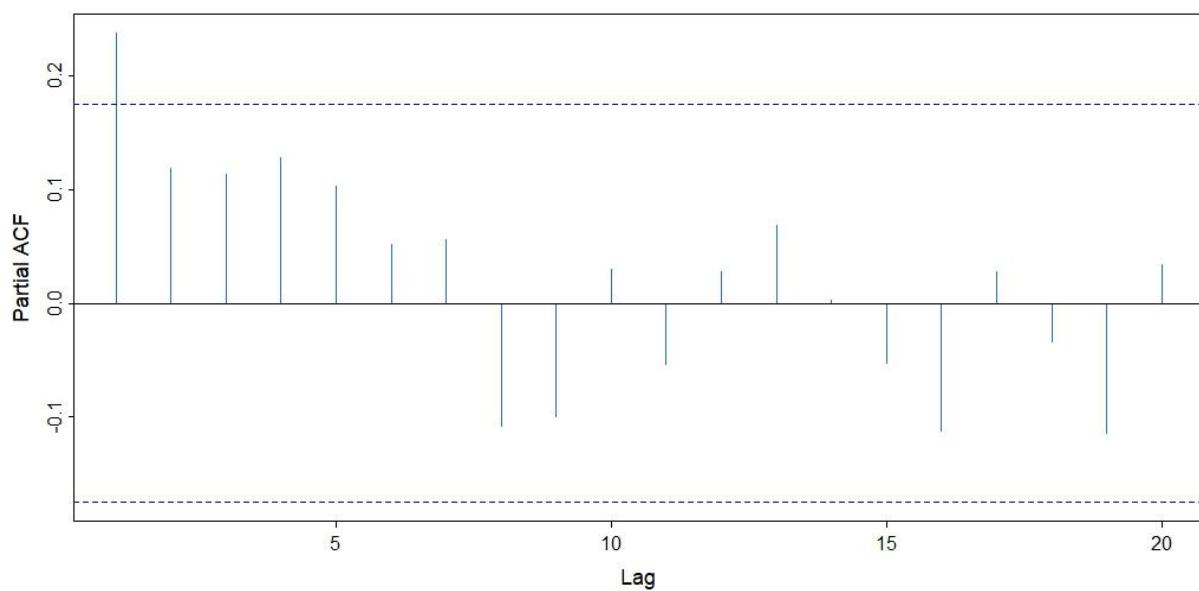
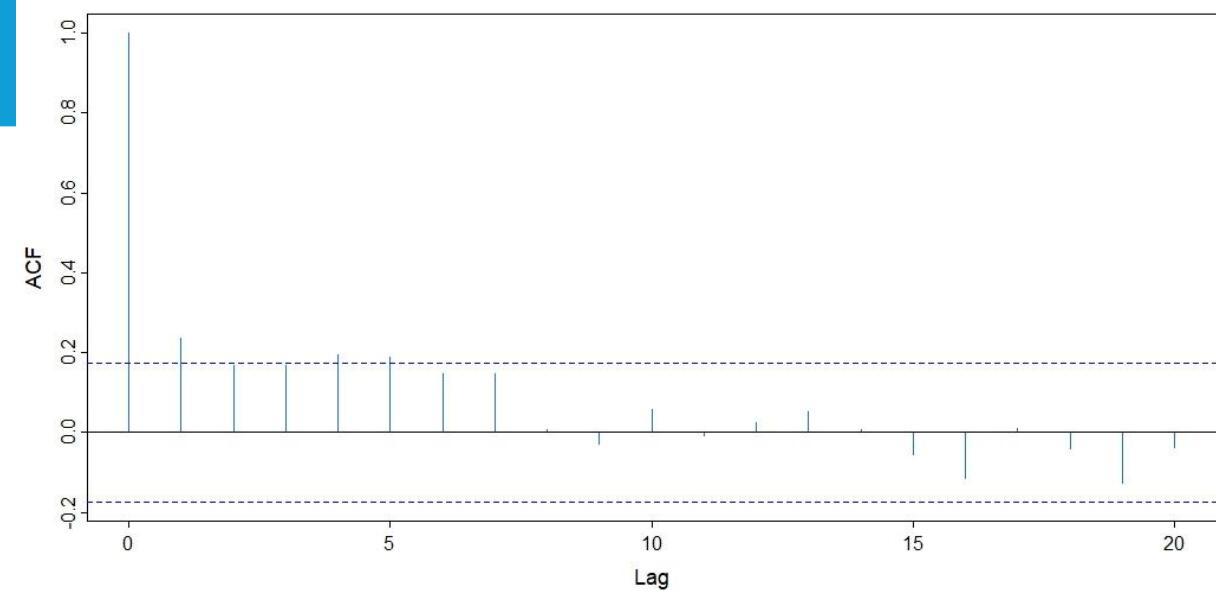
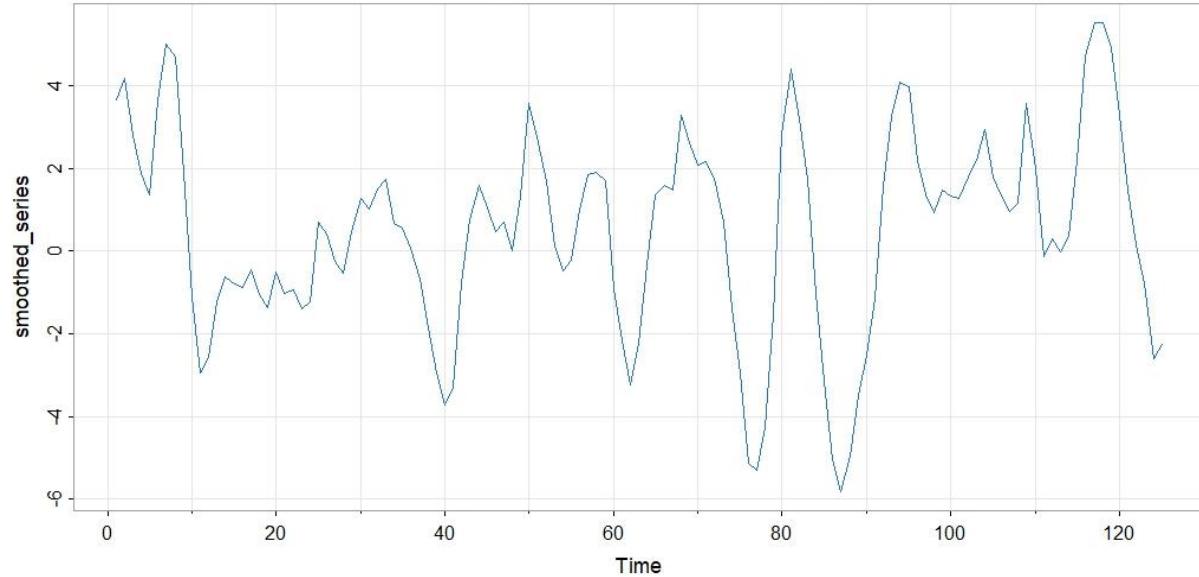


Weekly Vehicle Stop Diagnostics

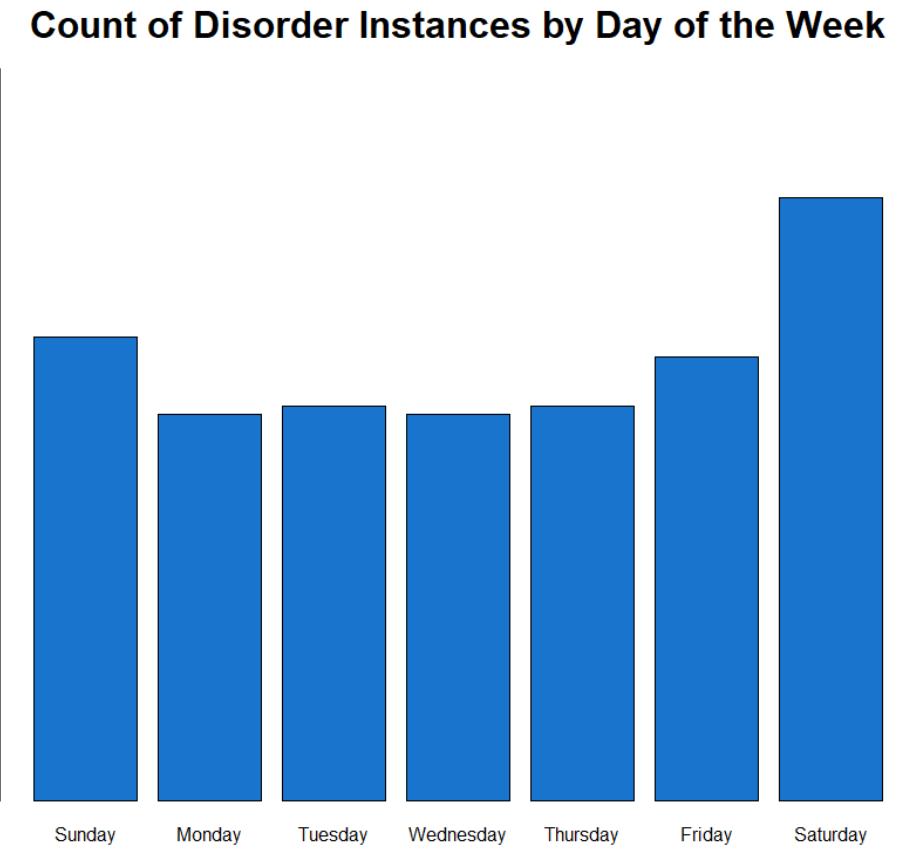
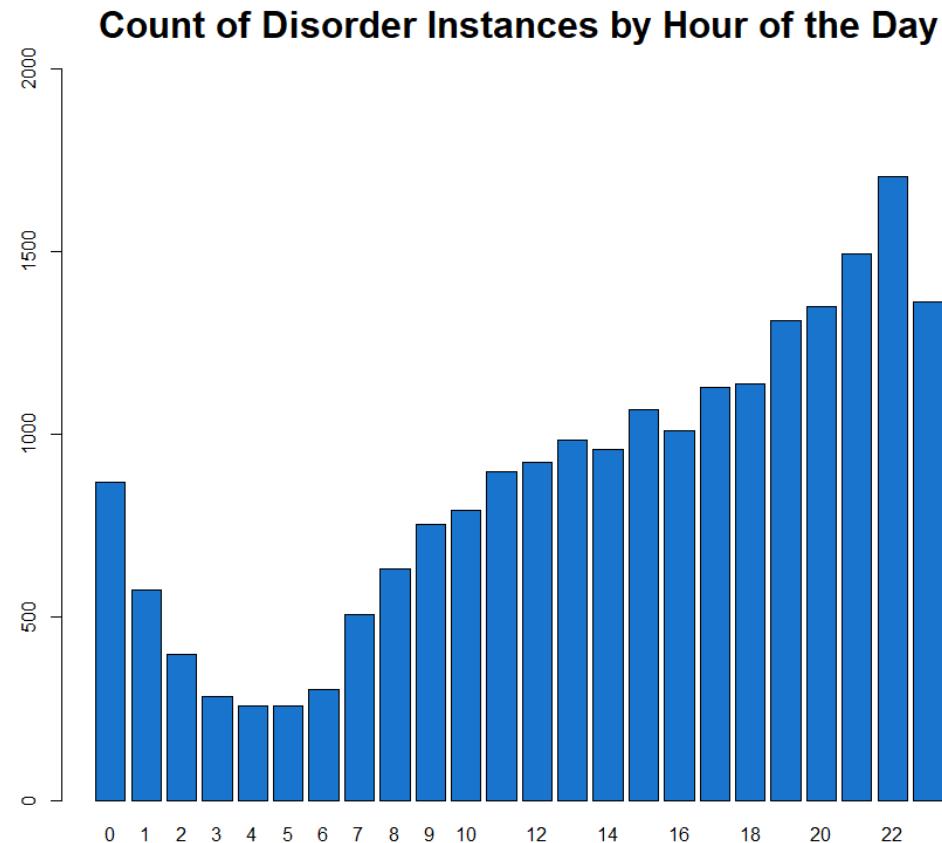
Vehicle Stop week Incidents



Vehicle Stop week Smoothed Autoregression

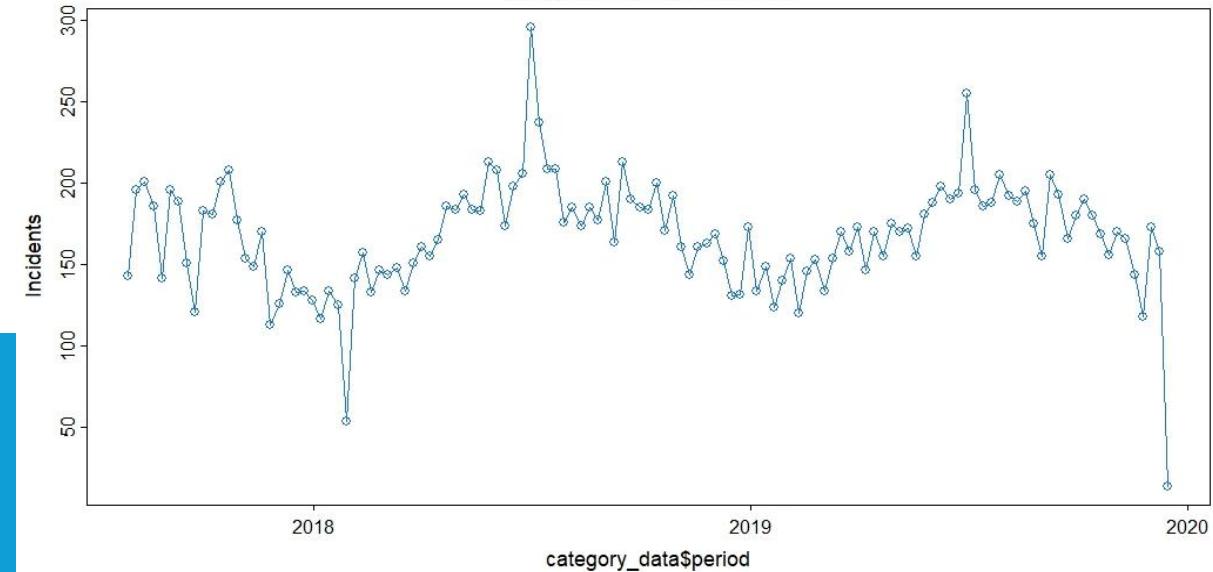


Disorder Instance Trends

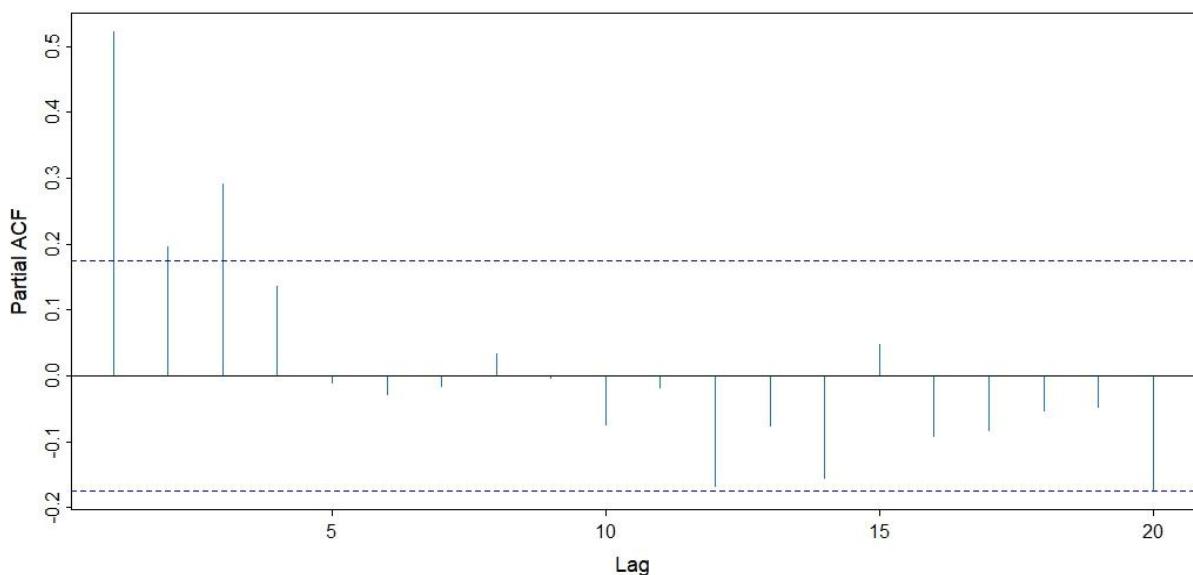
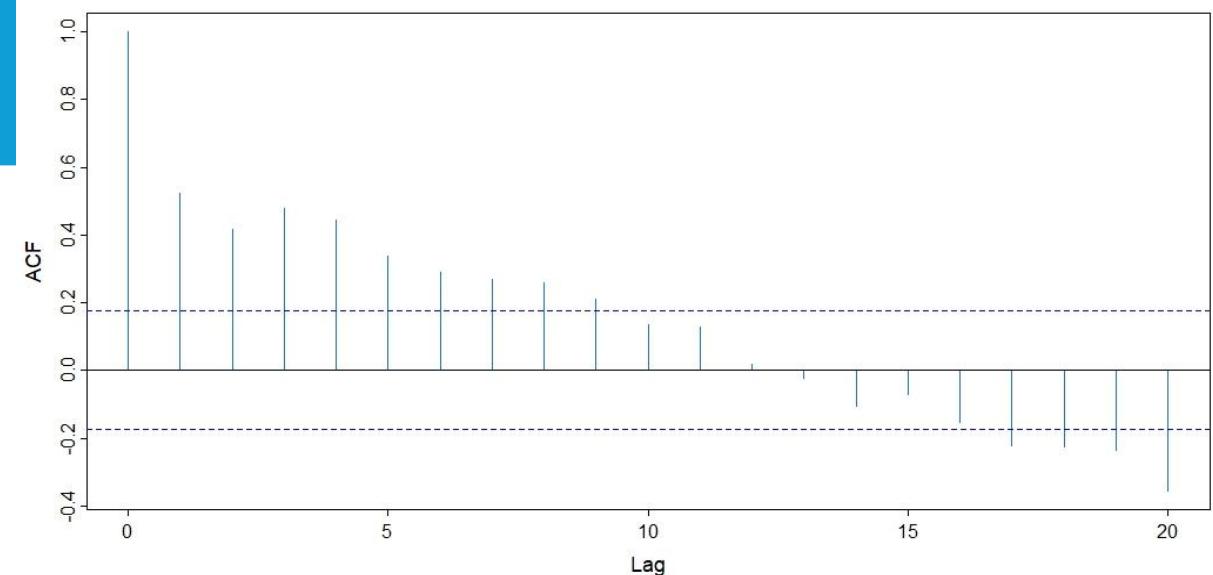
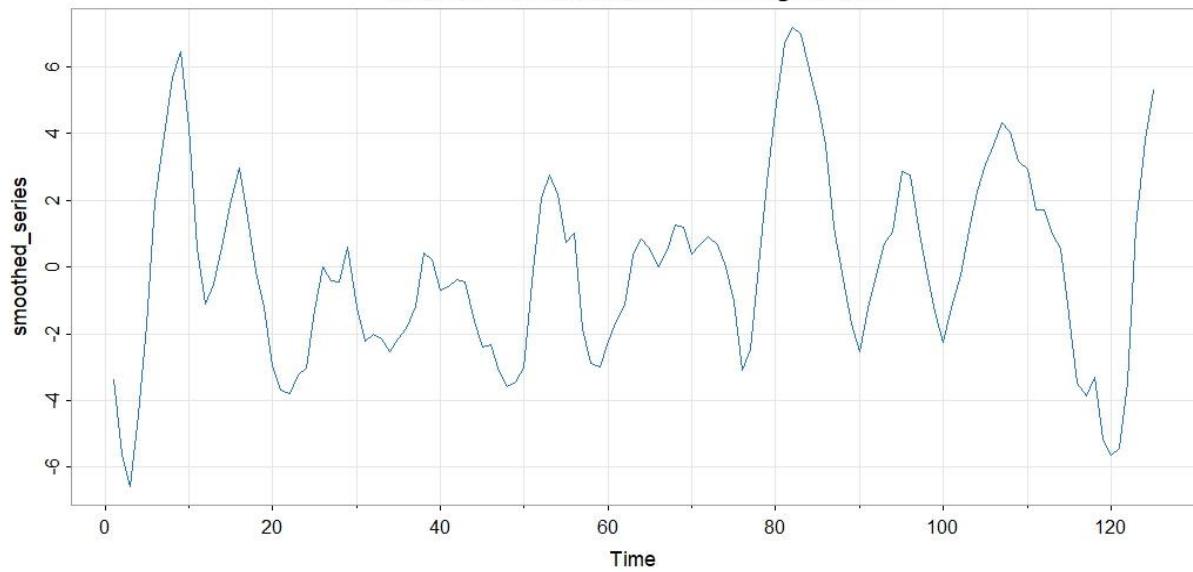


Disorder Diagnostics

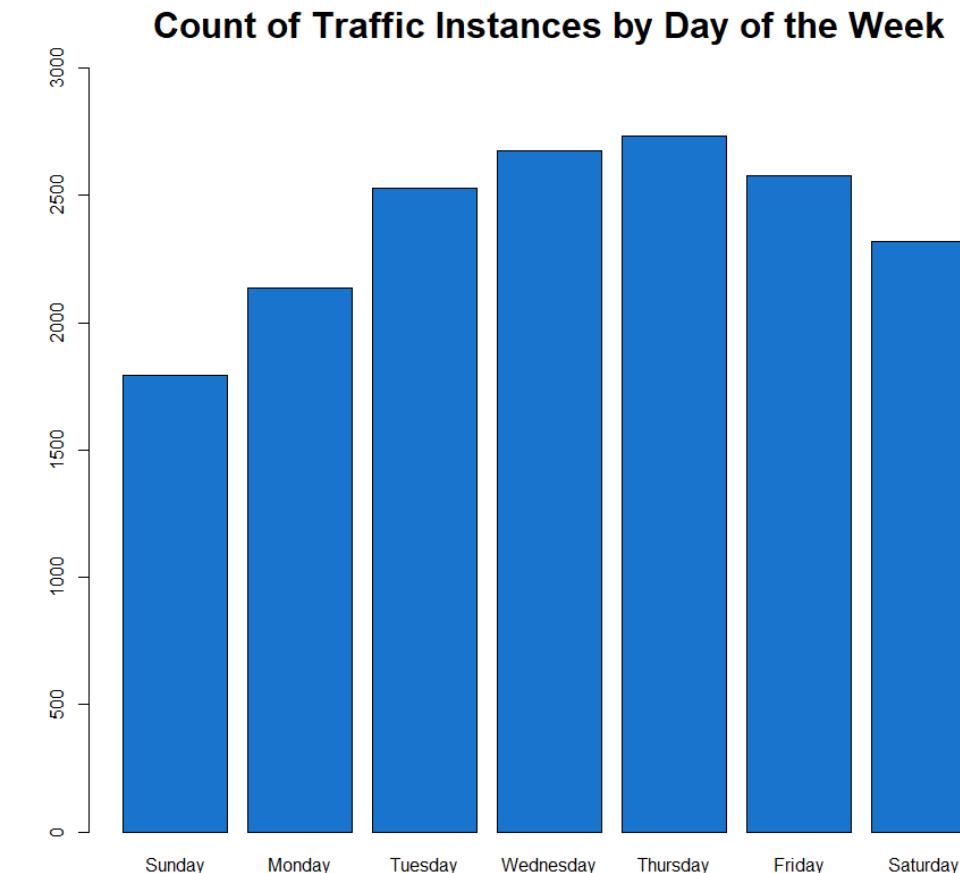
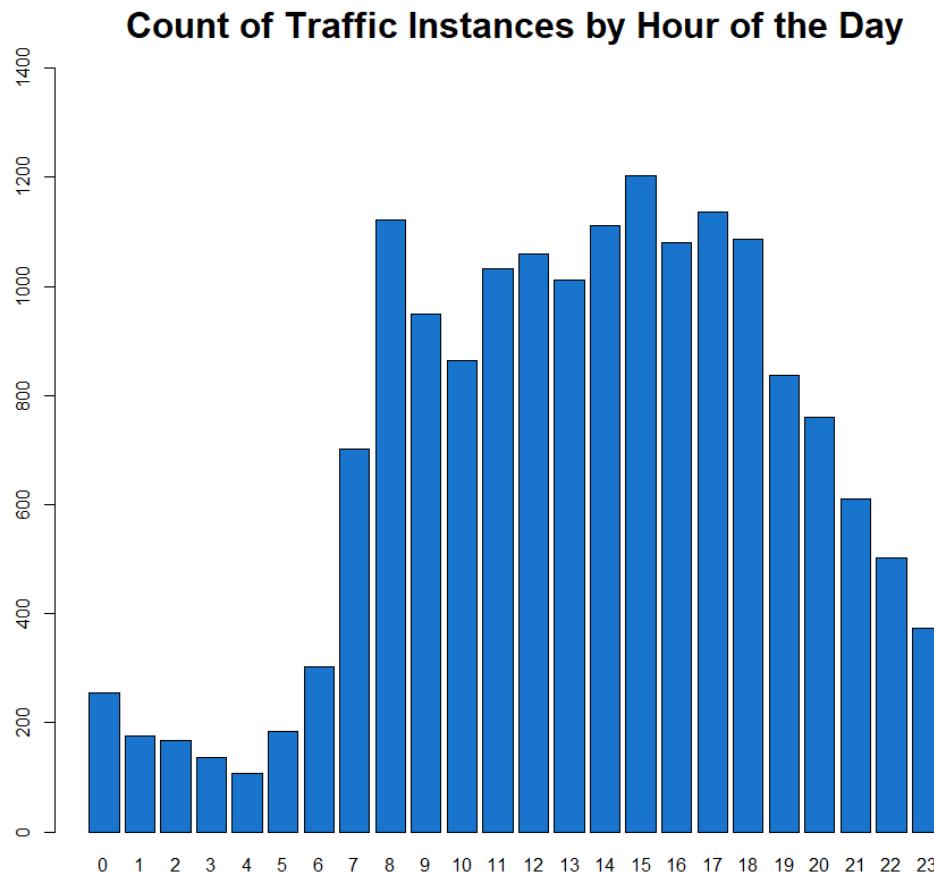
Disorder week Incidents



Disorder week Smoothed Autoregression

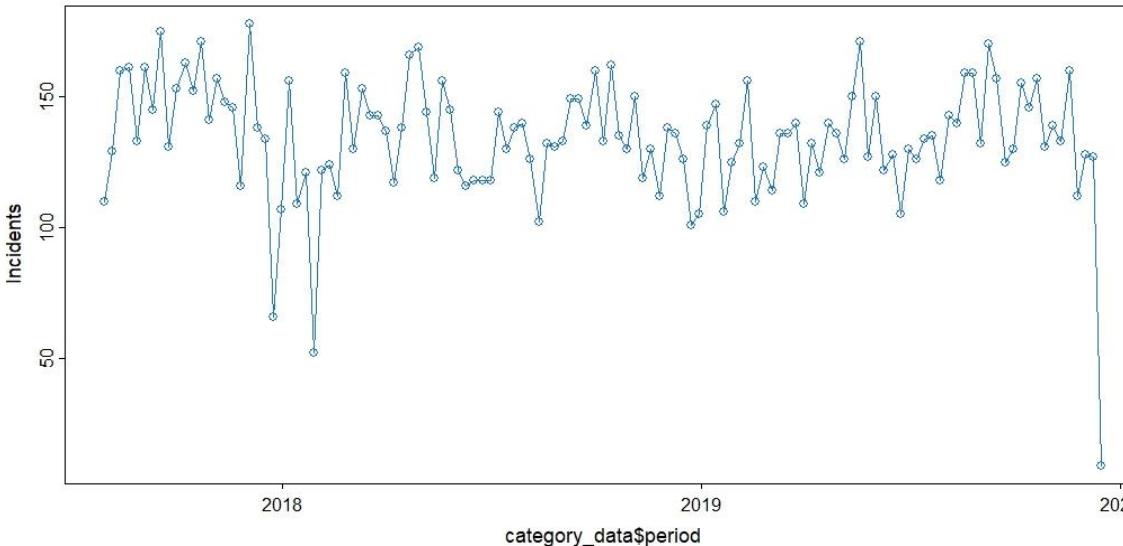


Traffic Instance Trends

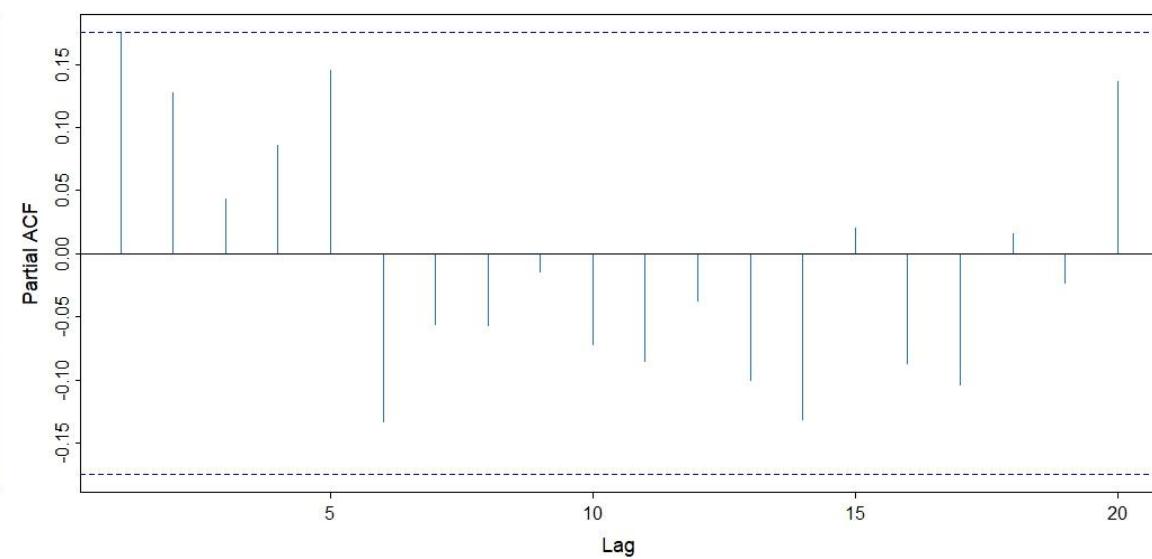
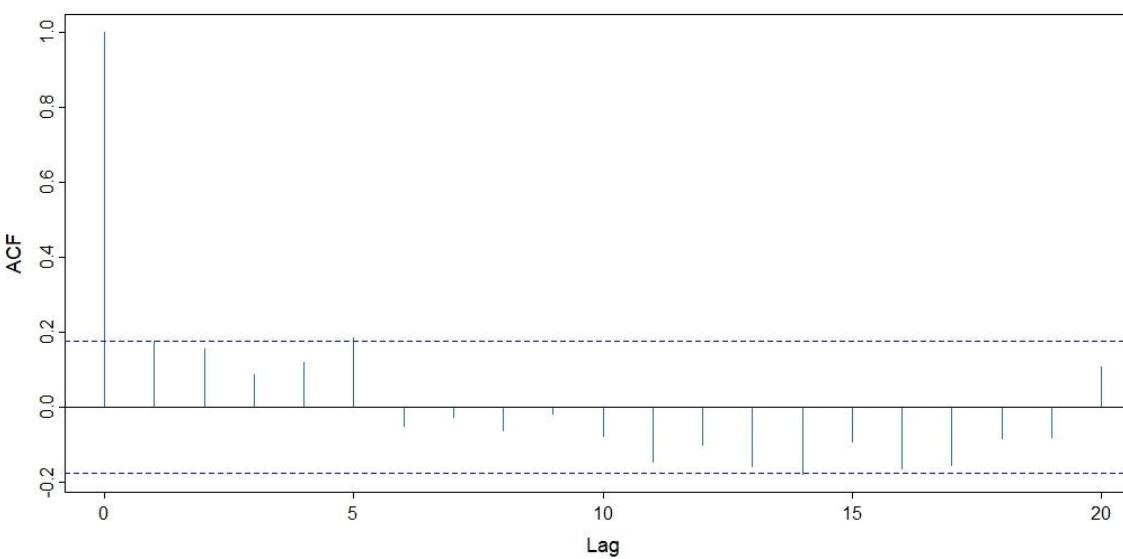
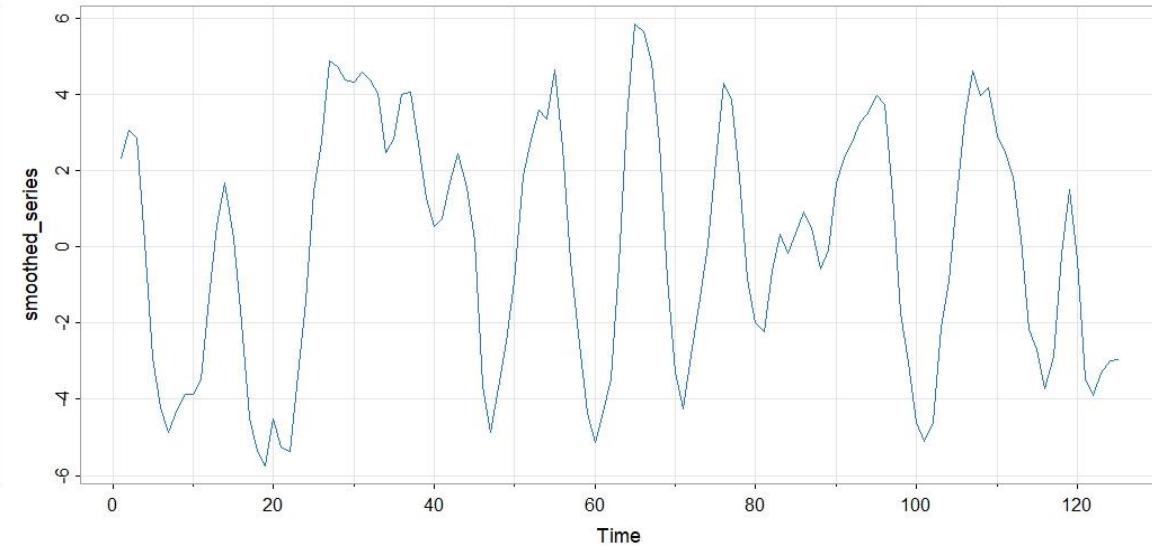


Traffic Diagnostics

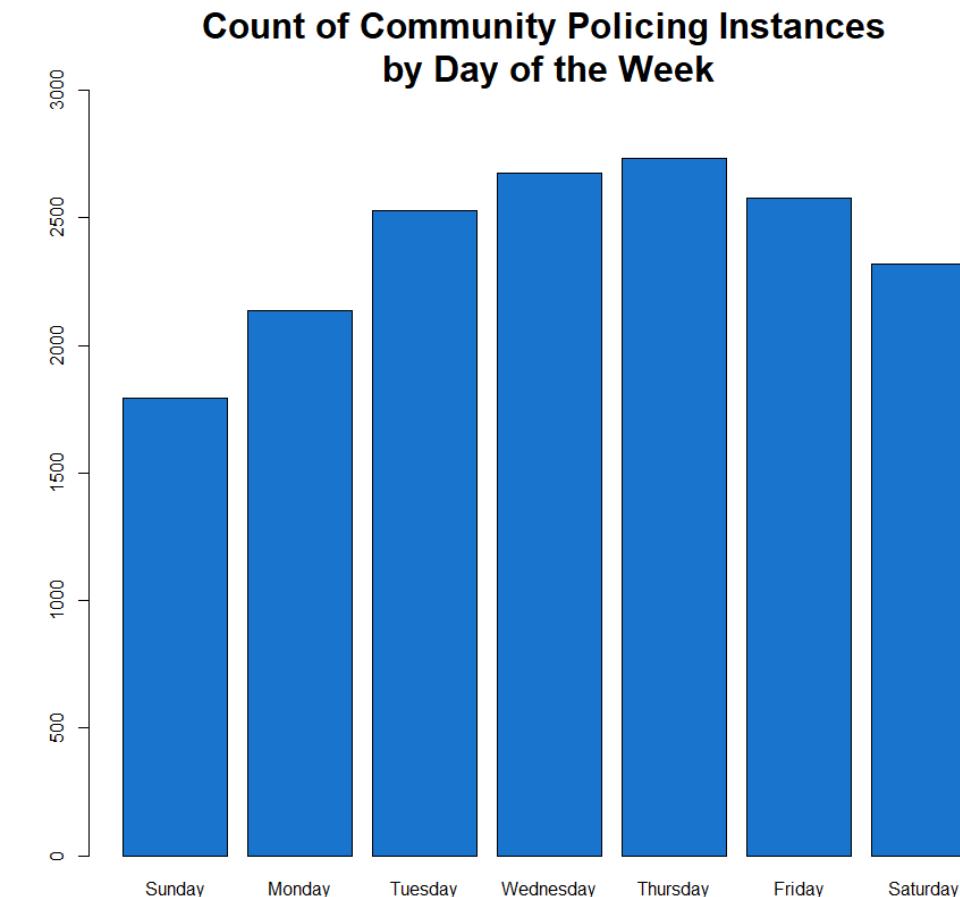
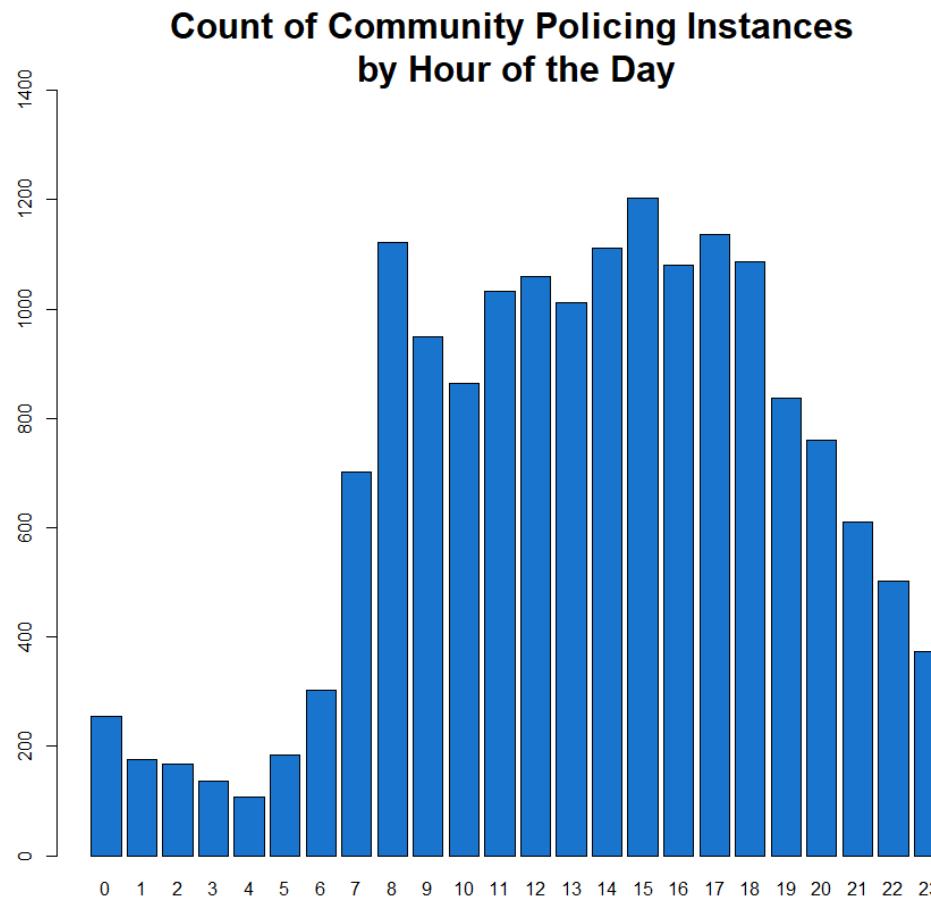
Traffic week Incidents



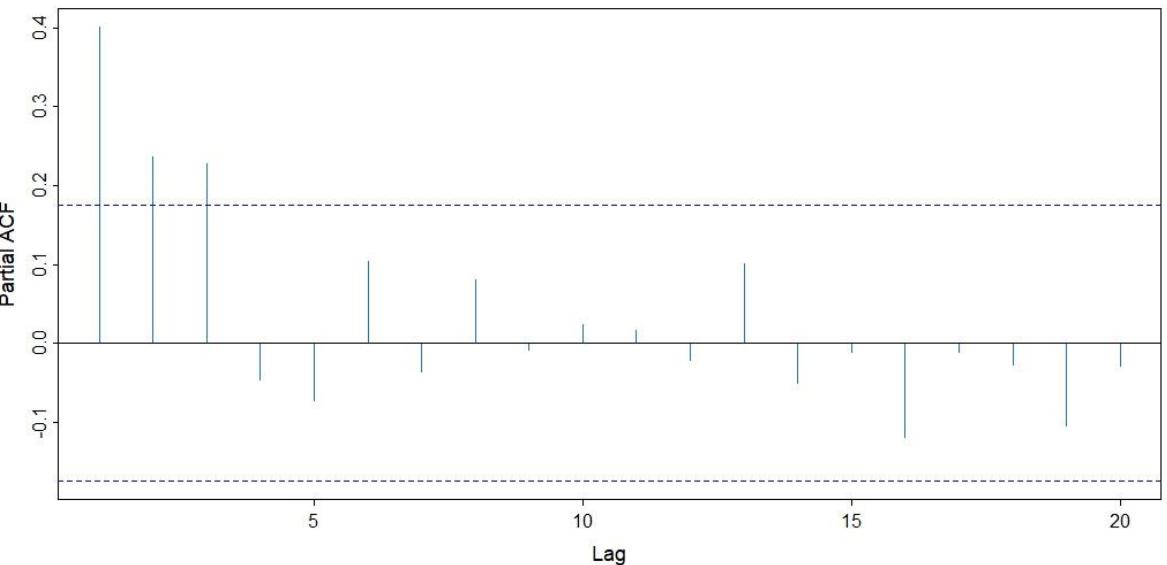
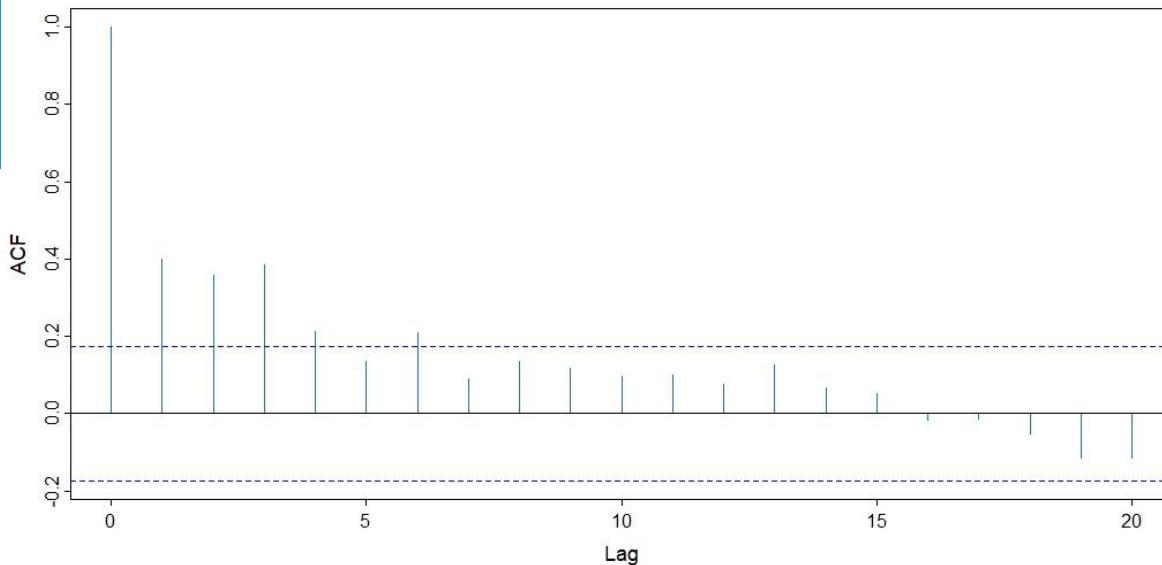
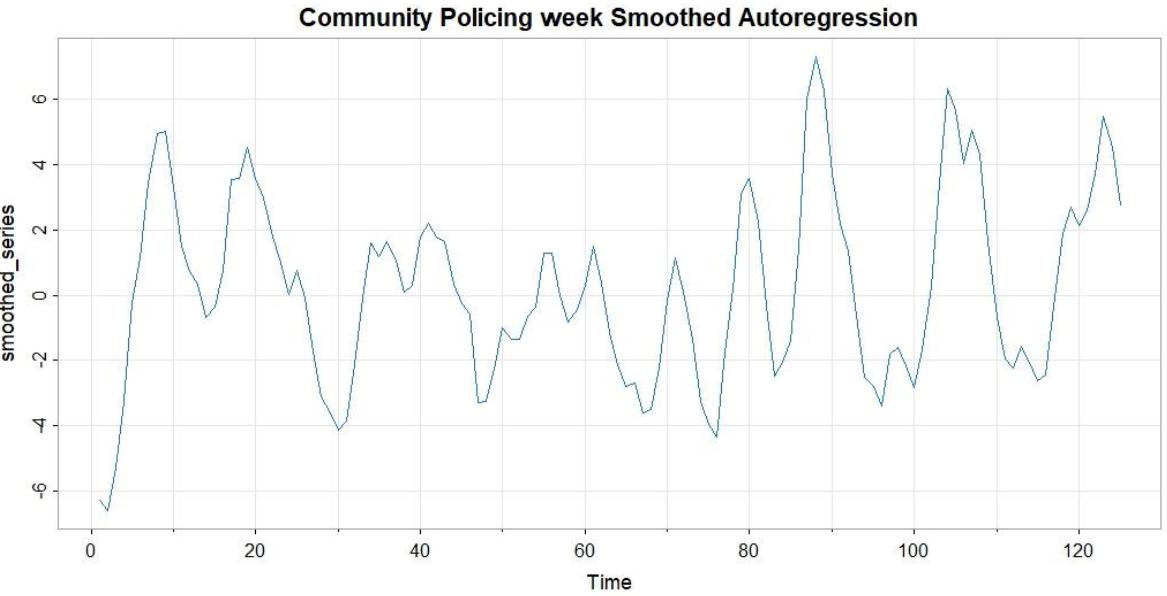
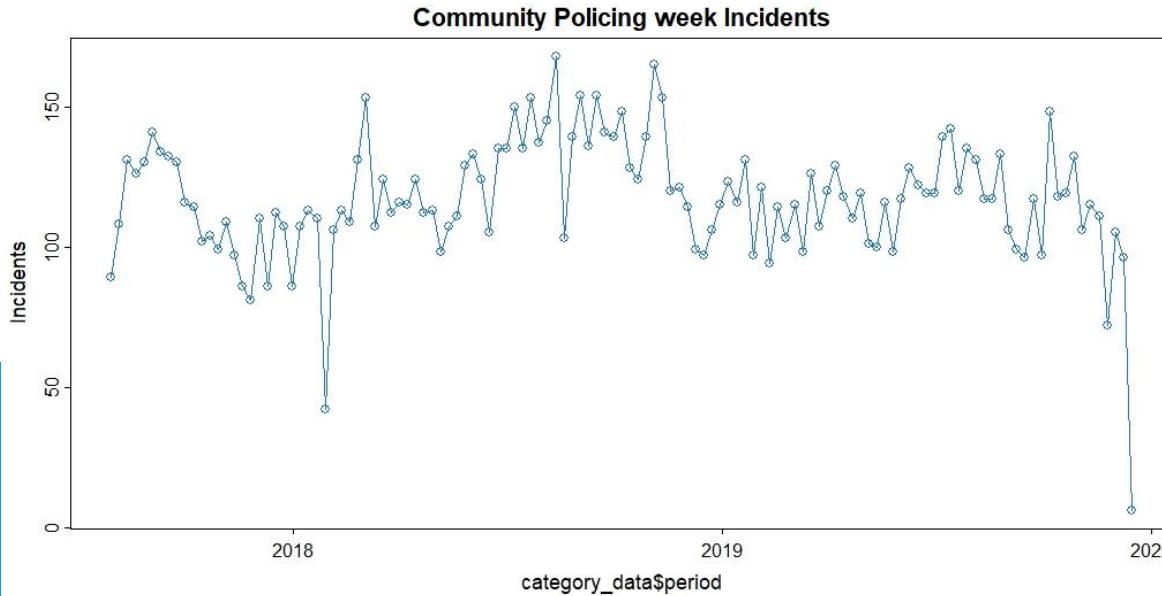
Traffic week Smoothed Autoregression



Community Policing Instance Trends



Community Policing Diagnostics



Santa Clara County Sheriff's Office

Daily Incidents

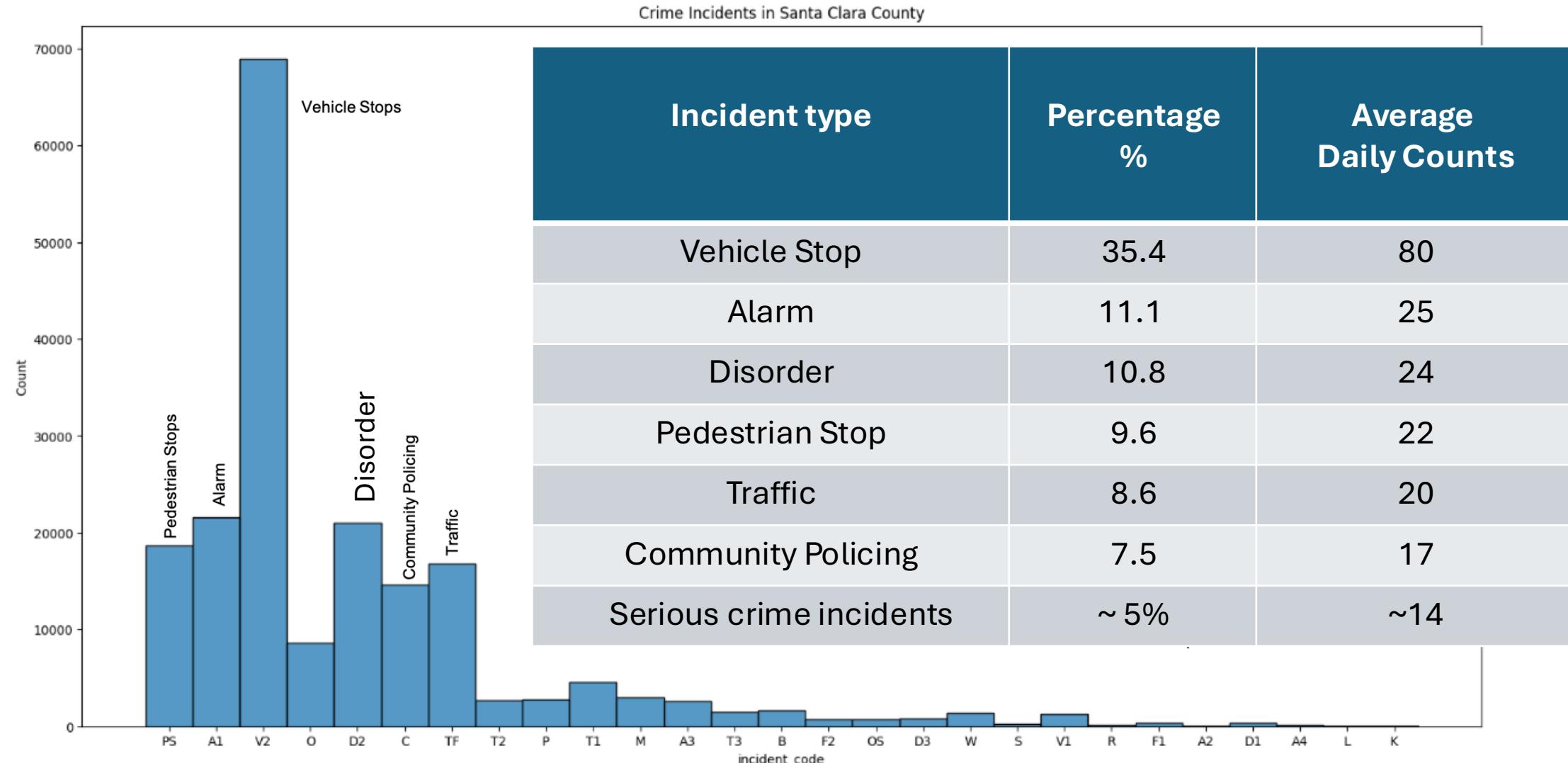
857 days of record

Day Average: 227 crime incidents

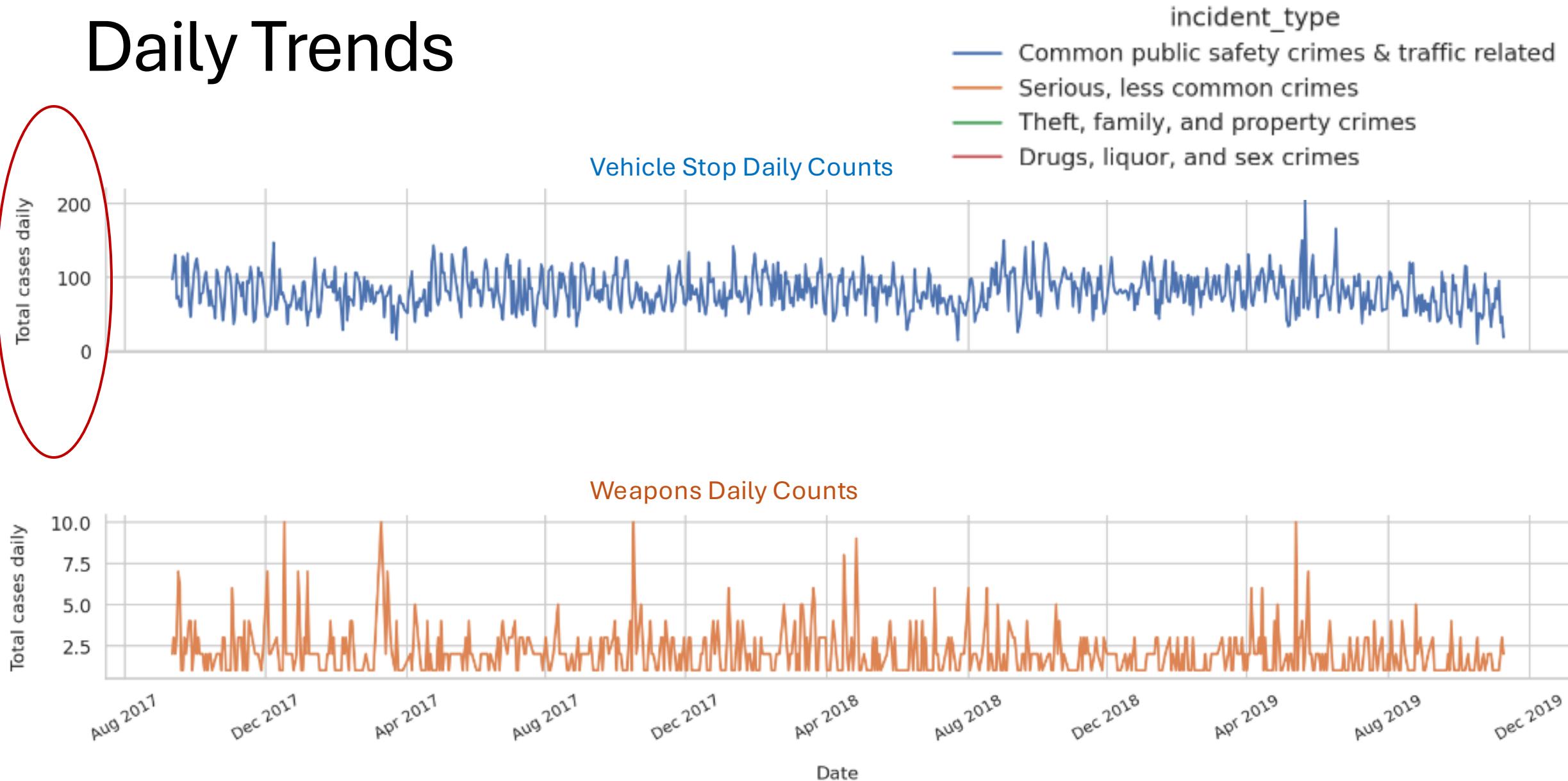
Missing Records Days	Day of week	Imputation Specific on Incident Type (counts in similar adjacent day)
2017-08-31	Thu	
2017-09-30	Sat	
2017-10-31	Tue	
2017-11-30	Thu	
2017-12-31	Sat	
2018-01-31	Wed	
2018-02-01	Thu	
2018-02-02	Fri	
2018-02-03	Sat	
2018-02-04	Sun	

Santa Clara County Sheriff's Office

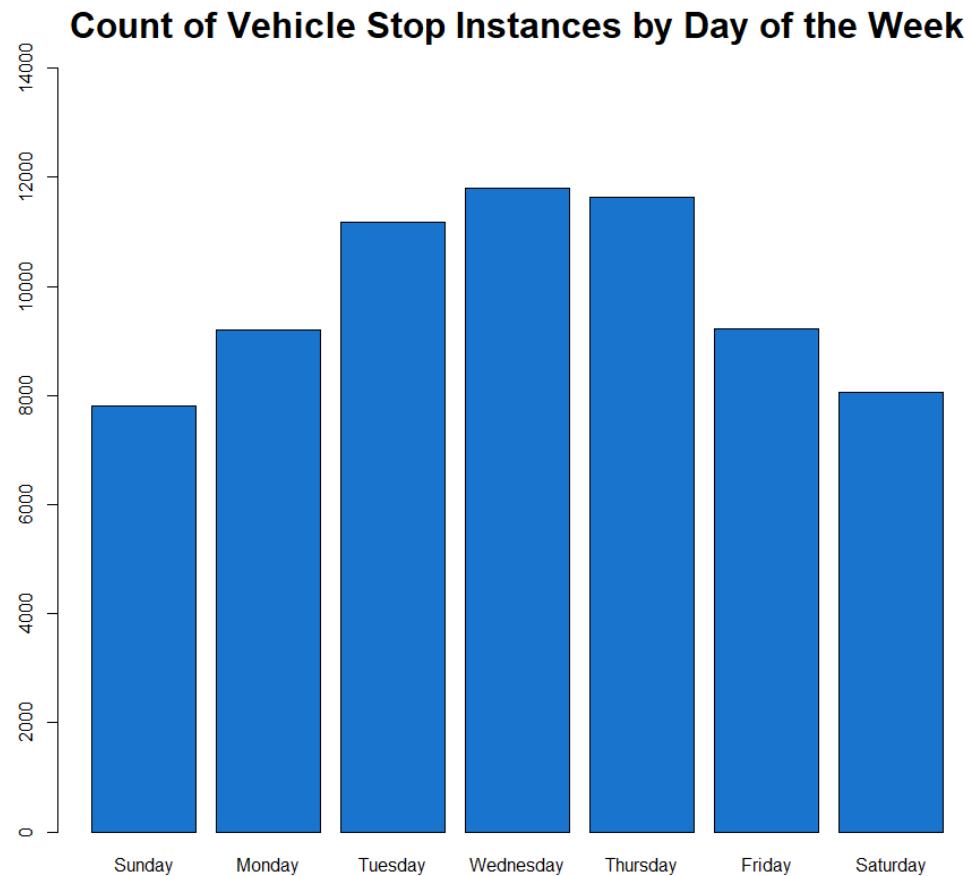
Daily Incidents



Daily Trends



Daily Trends: Vehicle Stop



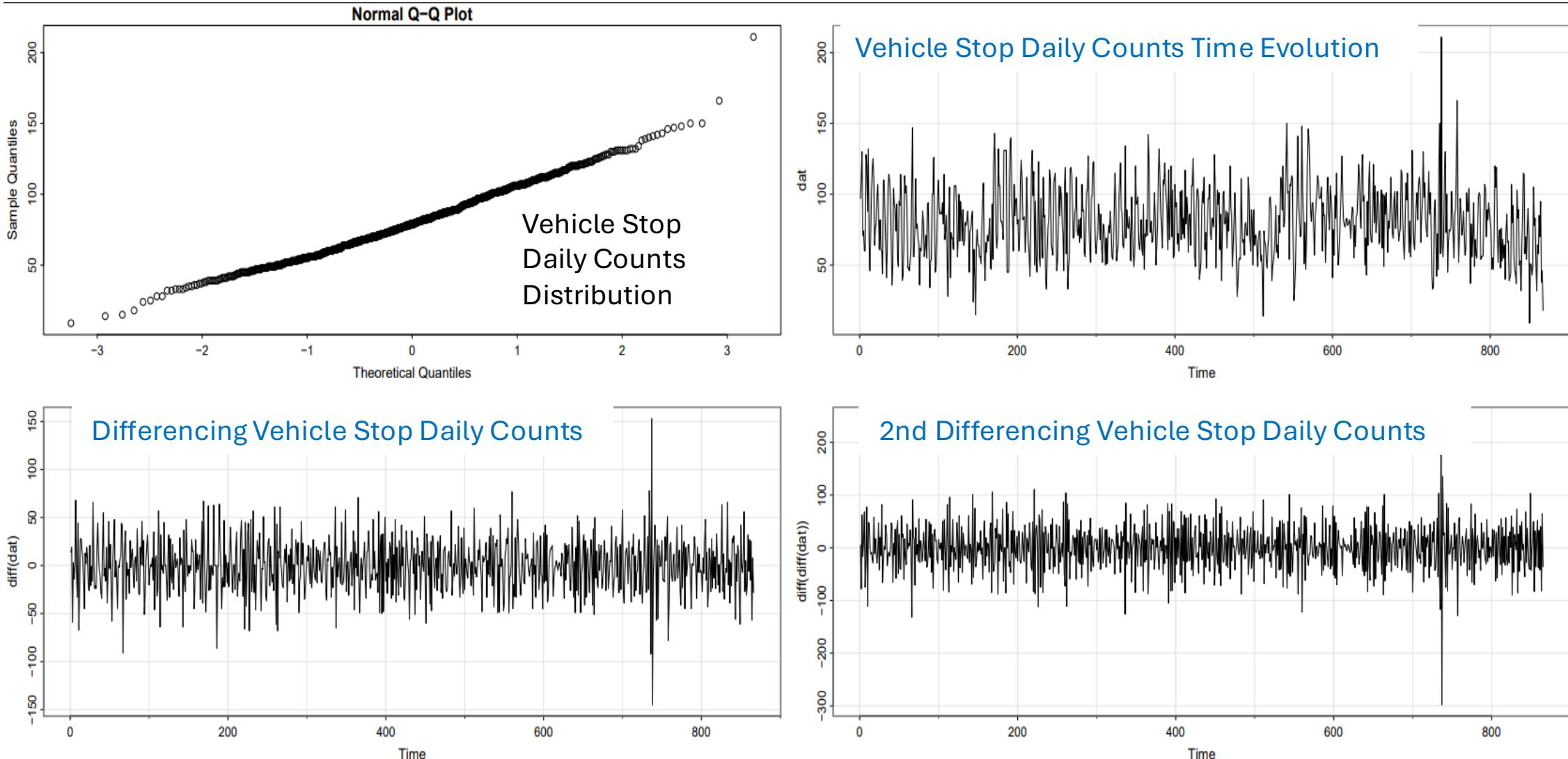
- Analysis Focus

on preventive policing incidents, each with more than 10 counts daily

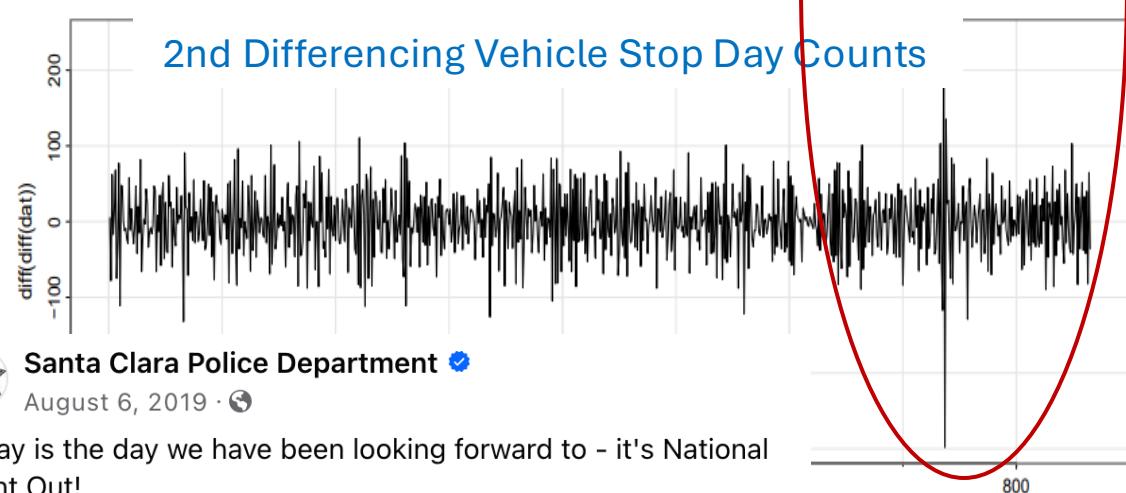
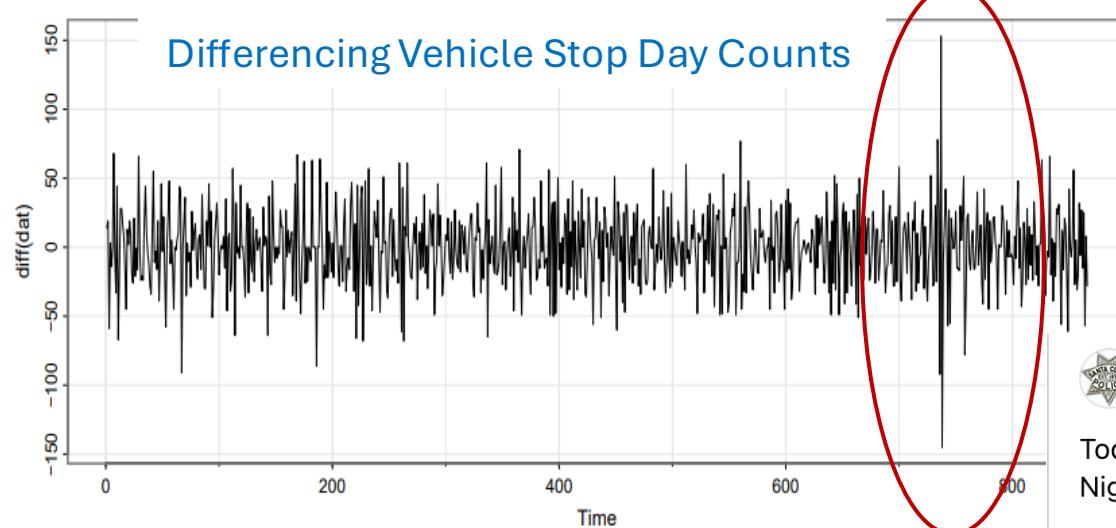
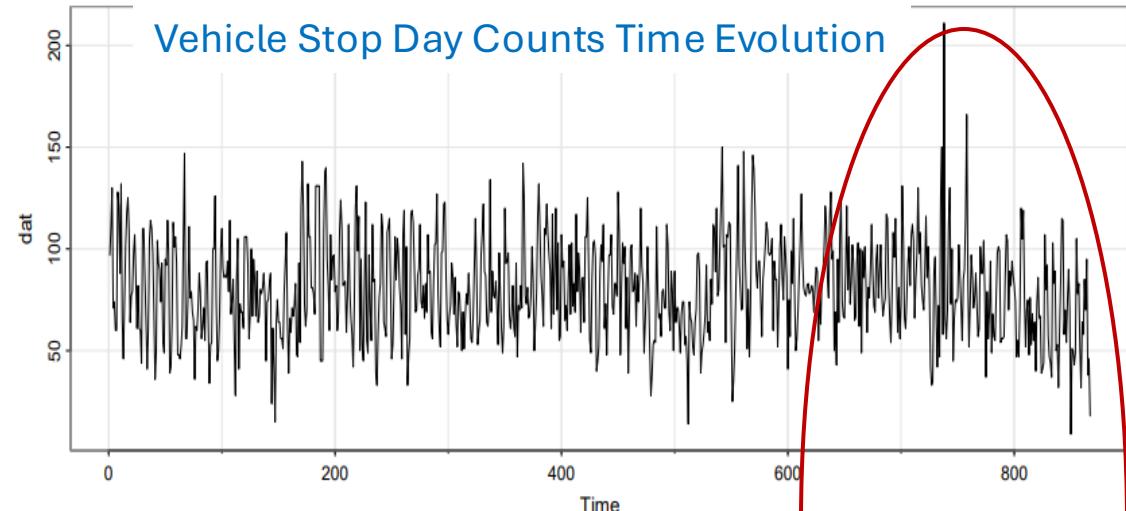
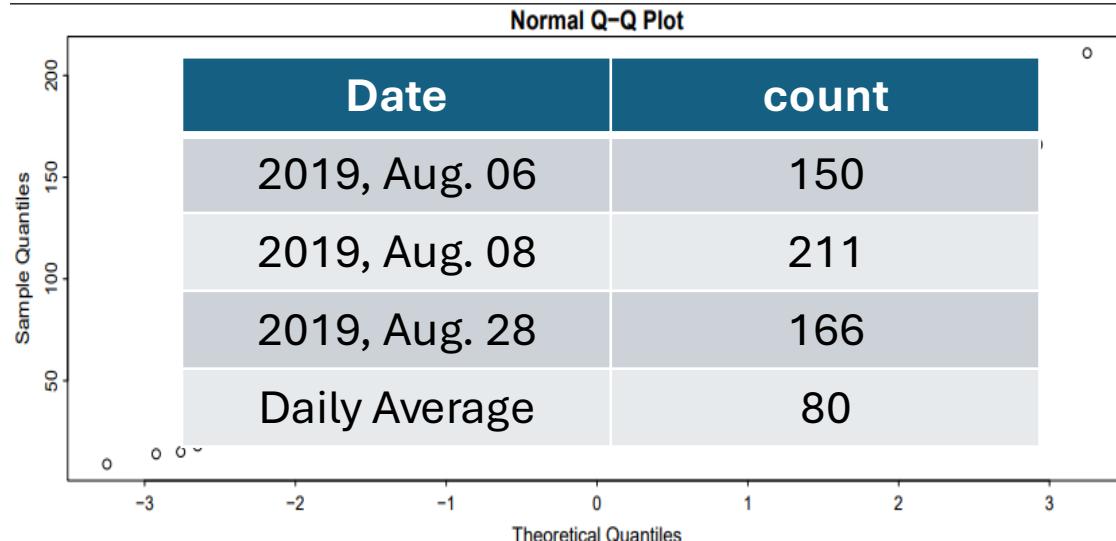
- Vehicle Stops - total 68,912
~ 35.4% of the total data

Daily Average: 80

Daily Vehicle Stop Diagnostics



Daily Vehicle Stop Diagnostics

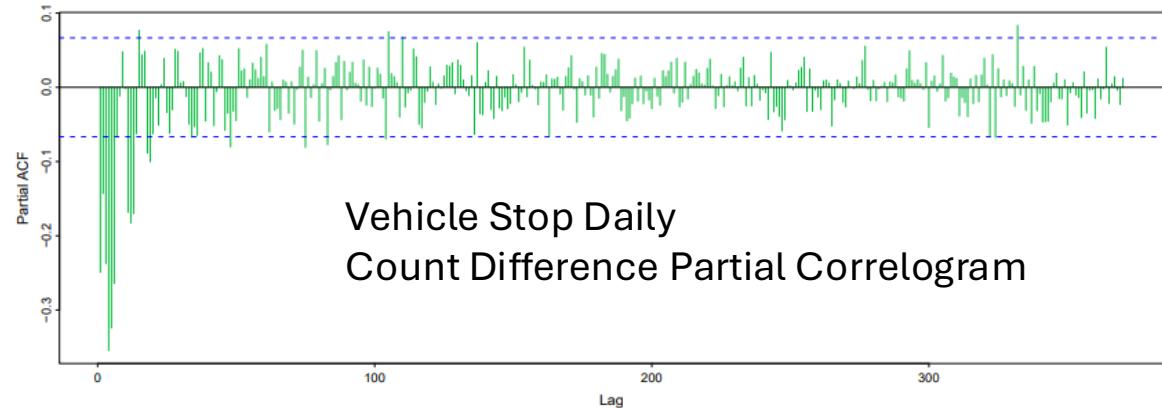
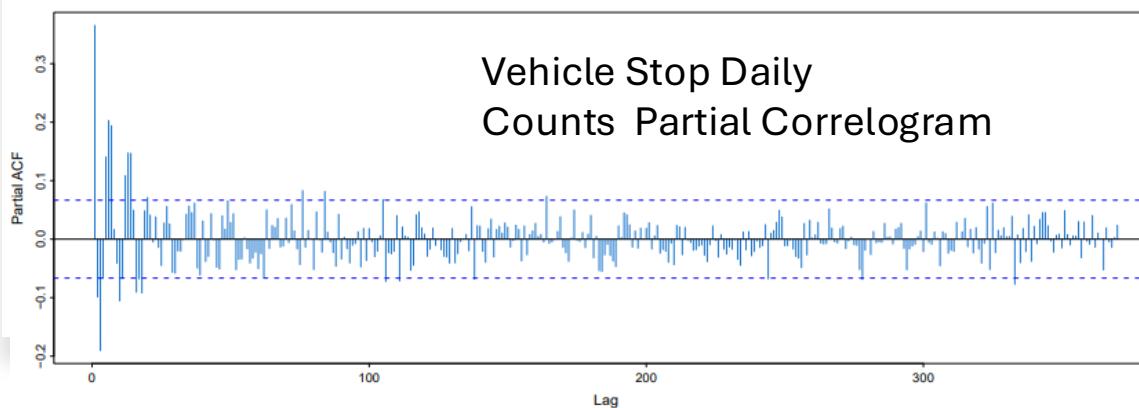
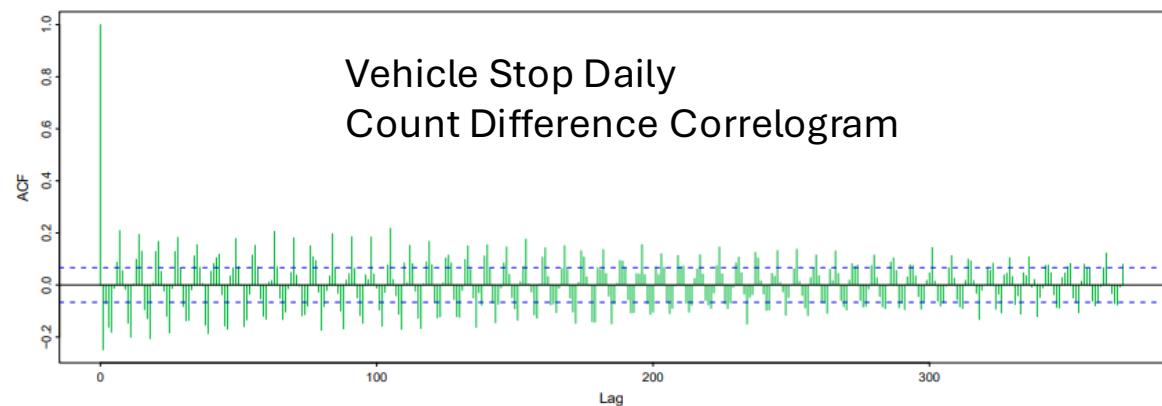
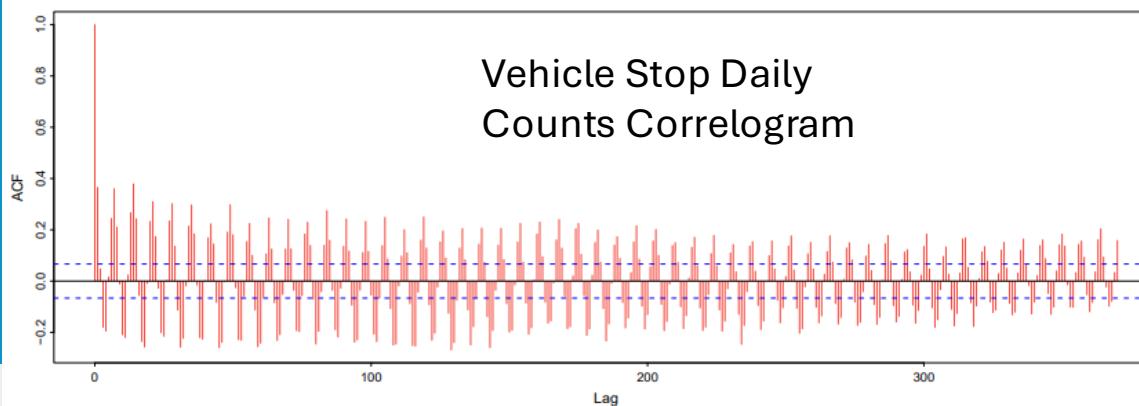
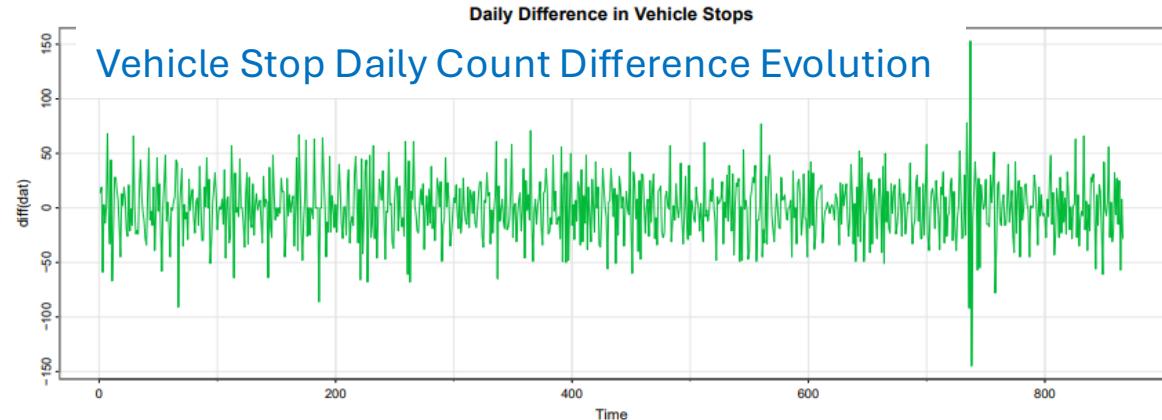
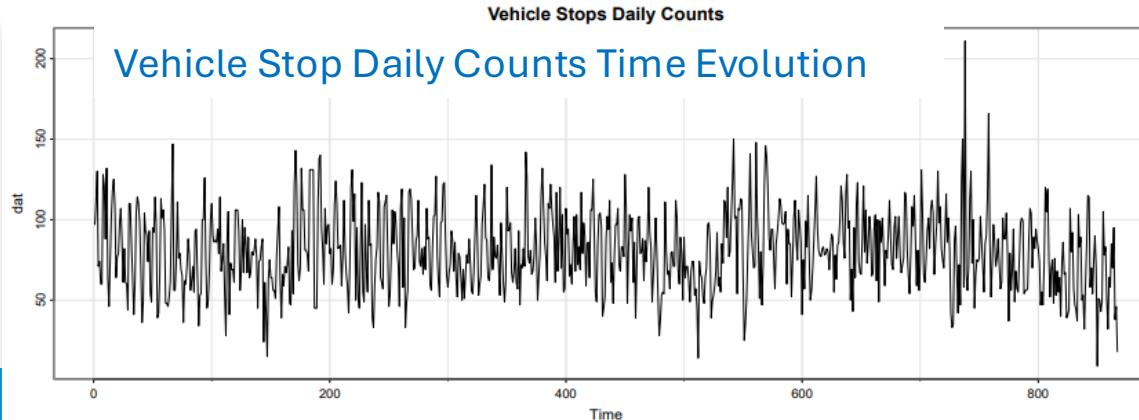


Santa Clara Police Department
August 6, 2019 ·

Today is the day we have been looking forward to - it's National Night Out!

Lock your doors, turn on your porch light and spend a beautiful evening outside with your neighbors.

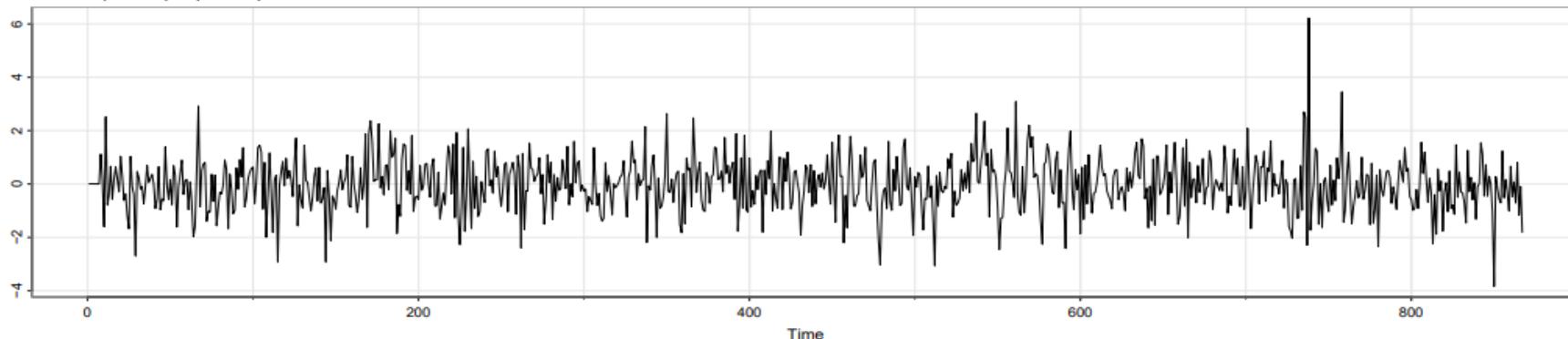
Daily Vehicle Stop Diagnostics



Daily Vehicle Stop Model and Goodness of Fit

Model: $(1, 0, 2) \times (0, 1, 1)_7$

Standardized Residuals



Vehicle Stop Day Count Model

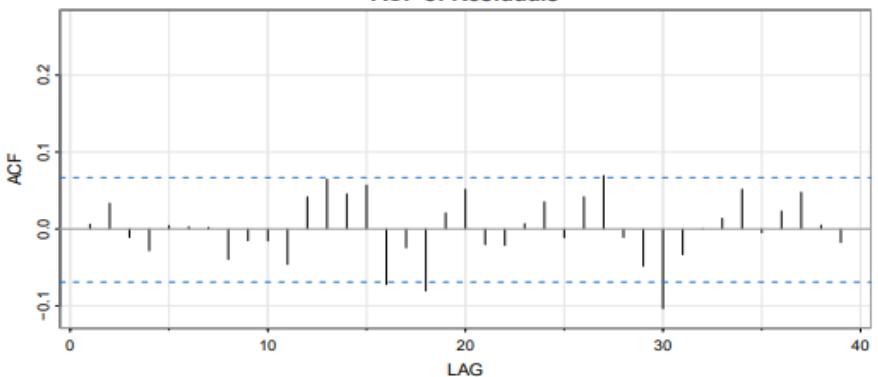
SARIMA

p=1, d=0, q=2

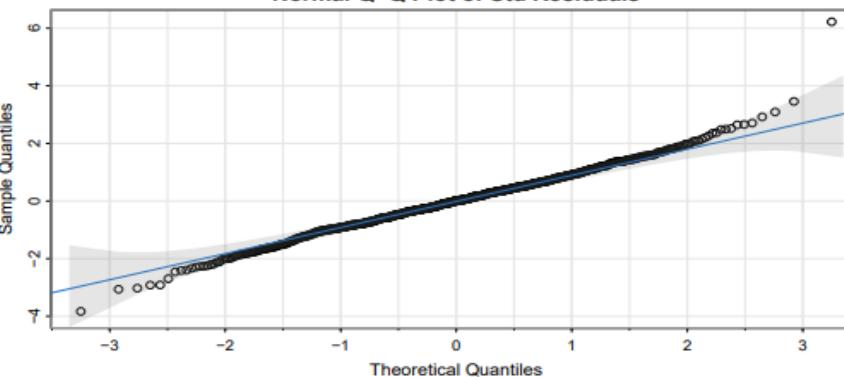
P=0, D=1, Q=1

7-day seasonality

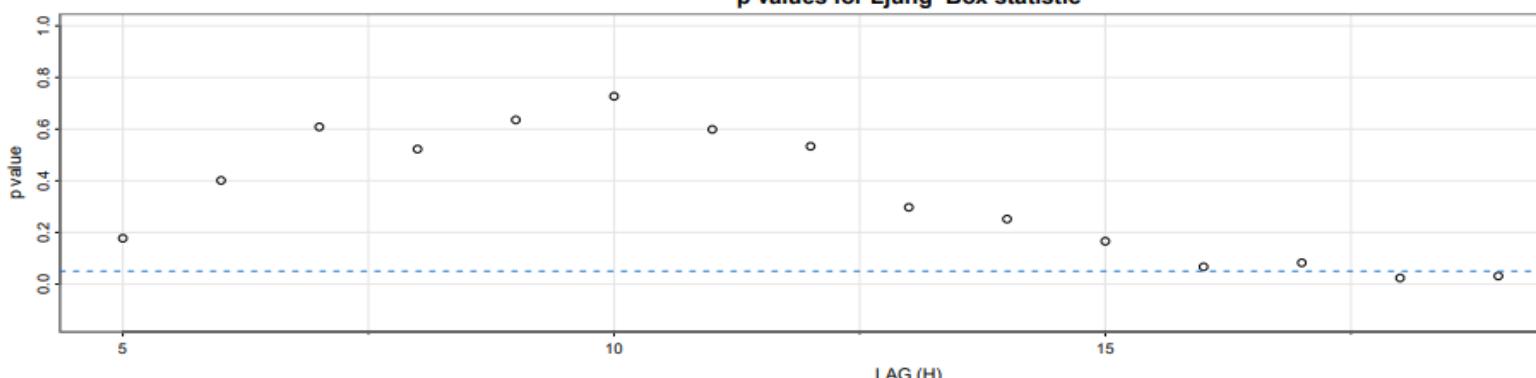
ACF of Residuals



Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic



Coefficients:

	Estimate	SE	t.value	p.value
ar1	0.9553	0.0292	32.7249	0.0000
ma1	-0.7295	0.0449	-16.2442	0.0000
ma2	-0.1470	0.0364	-4.0365	0.0001
sma1	-0.9512	0.0134	-70.9354	0.0000
constant	-0.0109	0.0160	-0.6792	0.4972

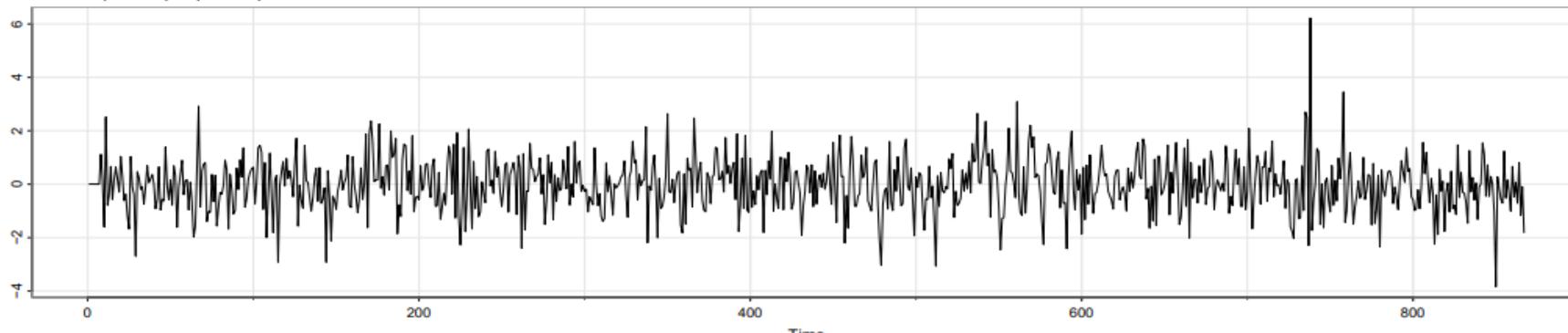
σ^2 estimated as 404.191 on 855 degrees of freedom

AIC = 8.871727 AICc = 8.871809 BIC = 8.904915

Daily Vehicle Stop Model and Goodness of Fit

Model: $(1, 0, 2) \times (0, 1, 1)_7$

Standardized Residuals



Vehicle Stop Day Count Model

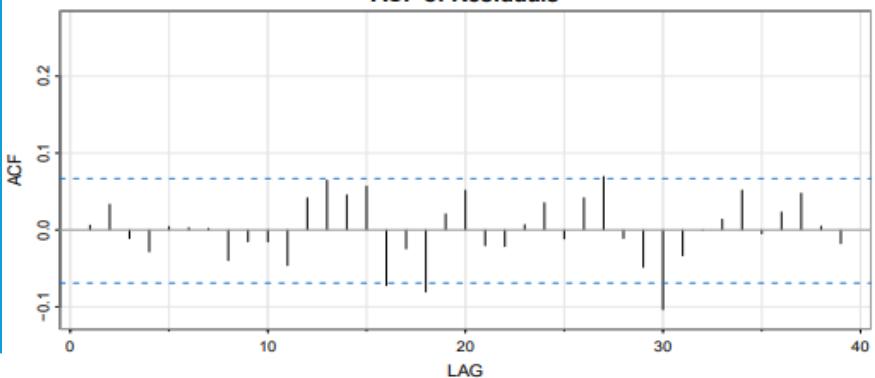
SARIMA

p=1, d=0, q=2

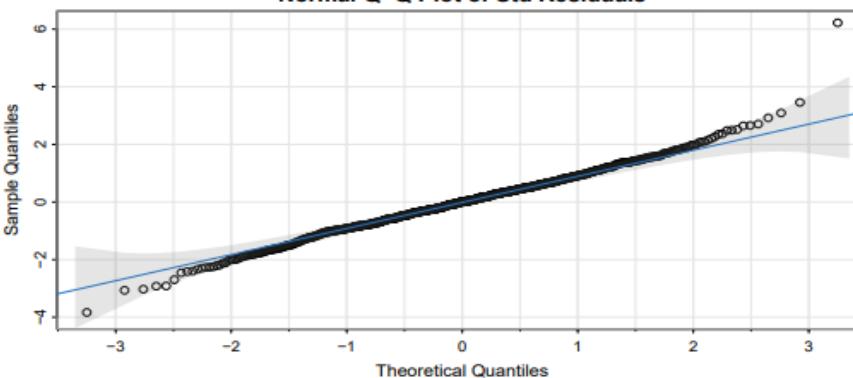
P=0, D=1, Q=1

7-day seasonality

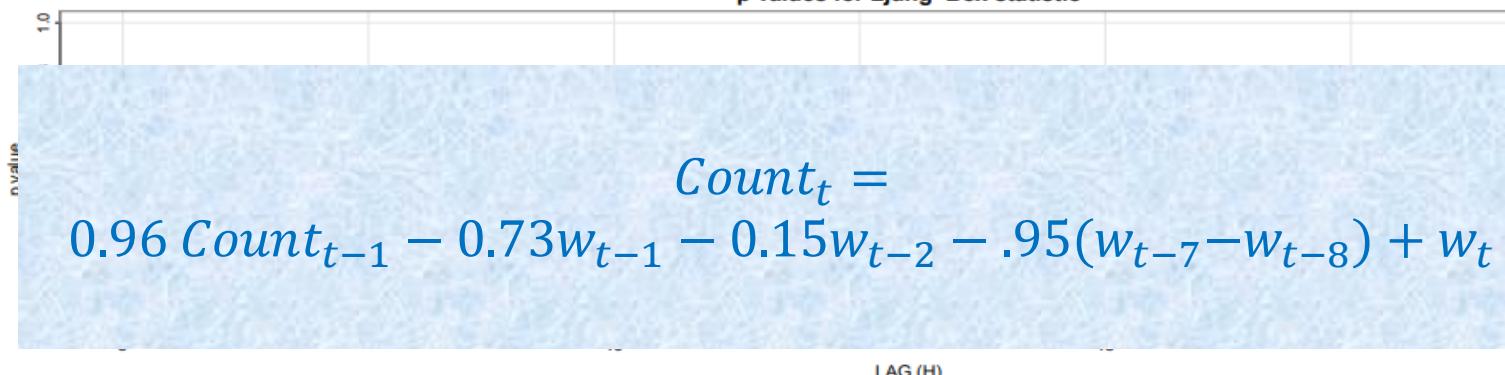
ACF of Residuals



Normal Q-Q Plot of Std Residuals



p values for Ljung-Box statistic



Coefficients:

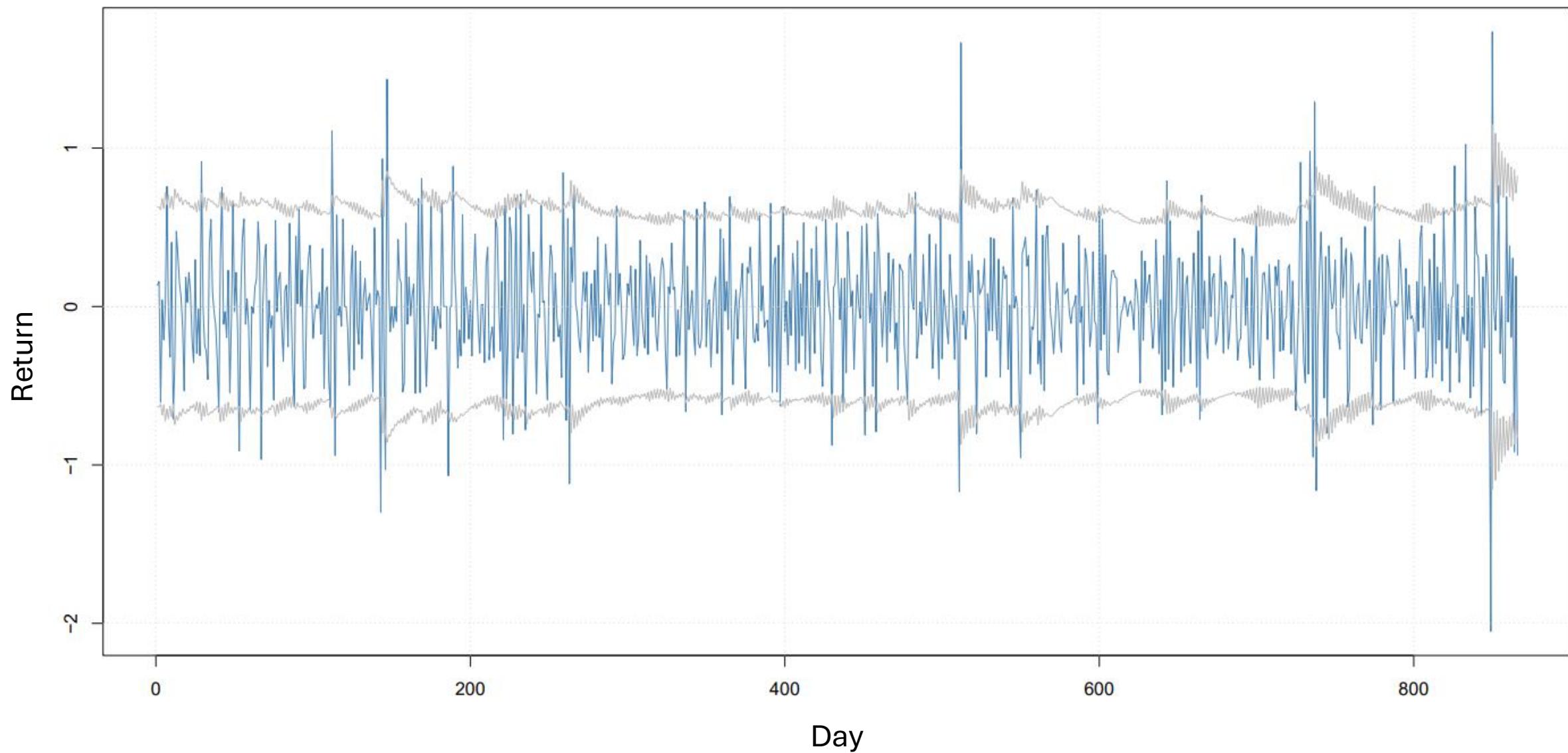
	Estimate	SE	t.value	p.value
ar1	0.9553	0.0292	32.7249	0.0000
ma1	-0.7295	0.0449	-16.2442	0.0000
ma2	-0.1470	0.0364	-4.0365	0.0001
sma1	-0.9512	0.0134	-70.9354	0.0000
constant	-0.0109	0.0160	-0.6792	0.4972

sigma^2 estimated as 404.191 on 855 degrees of freedom

AIC = 8.871727 AICc = 8.871809 BIC = 8.904915

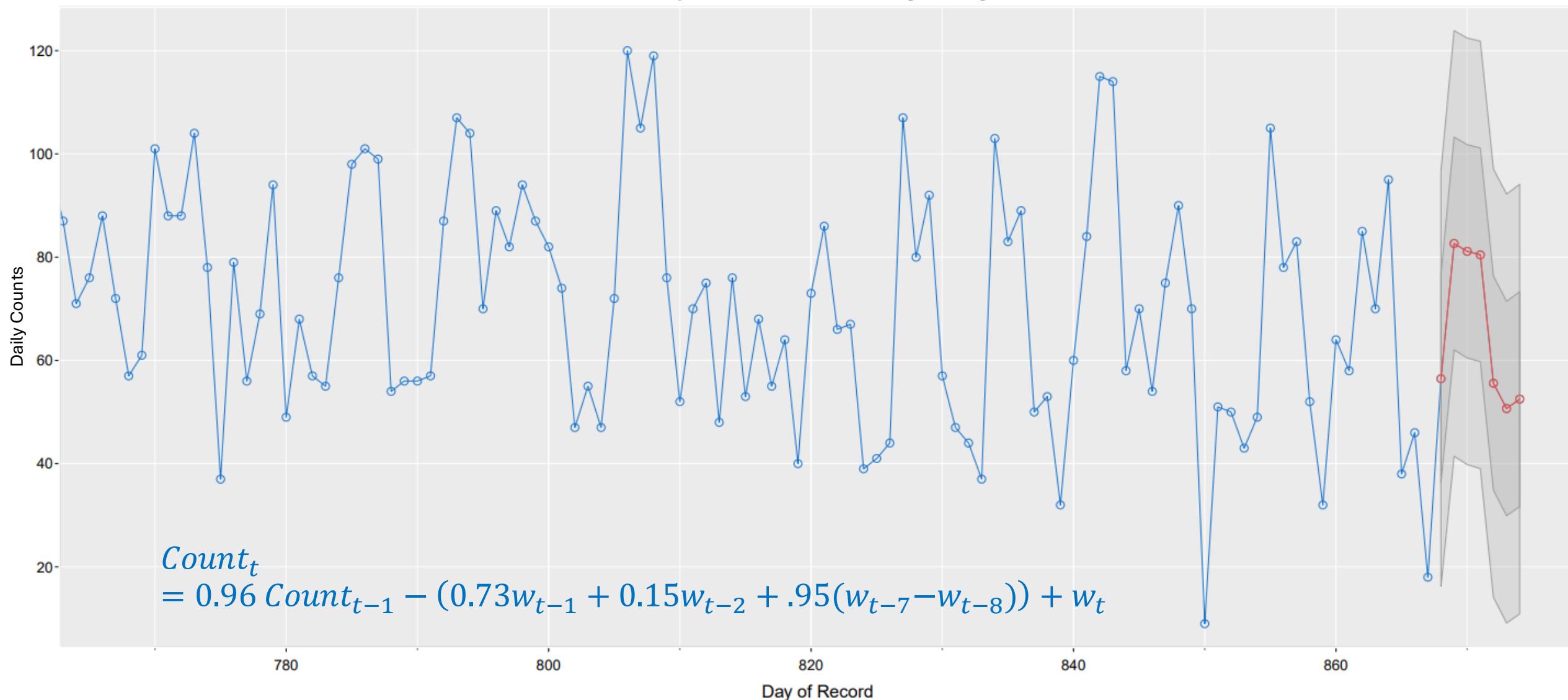
Daily Vehicle Stop Conditional Variance through GARCH

Series with 2 Conditional SD Superimposed

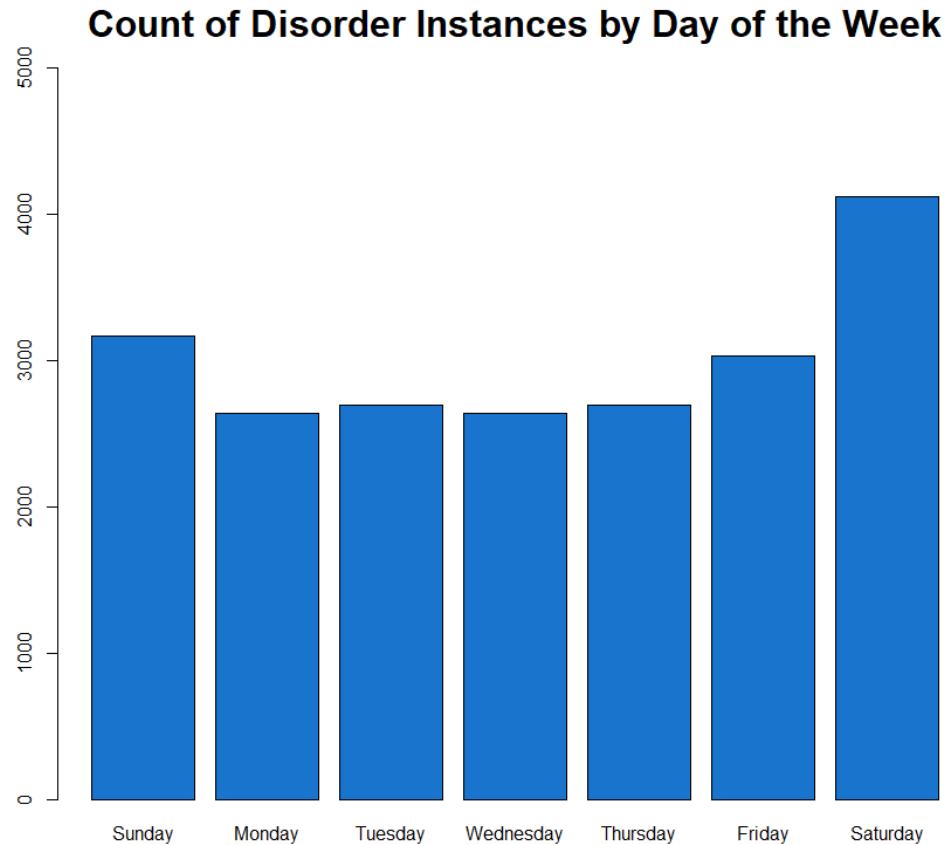


Daily Vehicle Stop Seven-Day Prediction for Dec. 2019

Vehicle Stops in Santa Clara County – Daily Counts



Daily Trends: Disorder

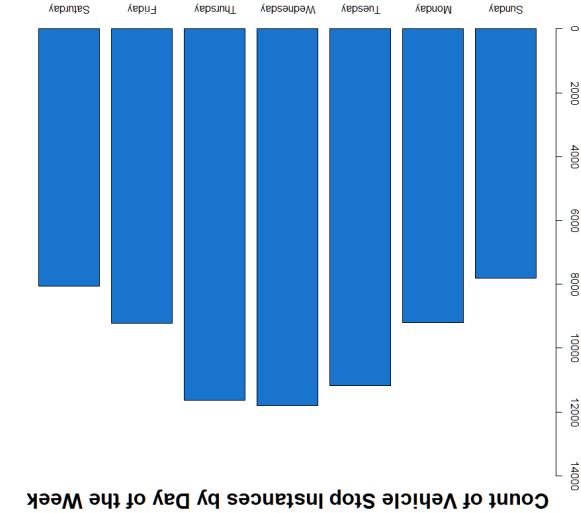


- **Analysis Focus**
on preventive policing incidents, each
with more than 10 counts daily
 - Disorder - total recorded 20,981
~10.8% of the total data

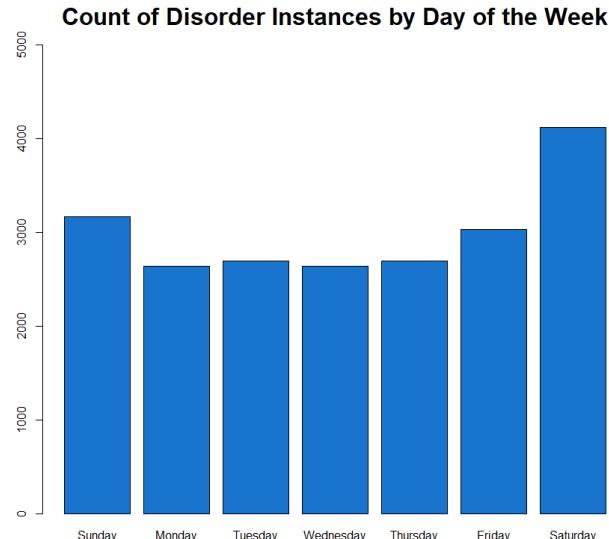
Daily Average: 24

Daily Trends: Disorder

Weekly pattern:
Vehicle Stop



Weekly pattern:
Disorder

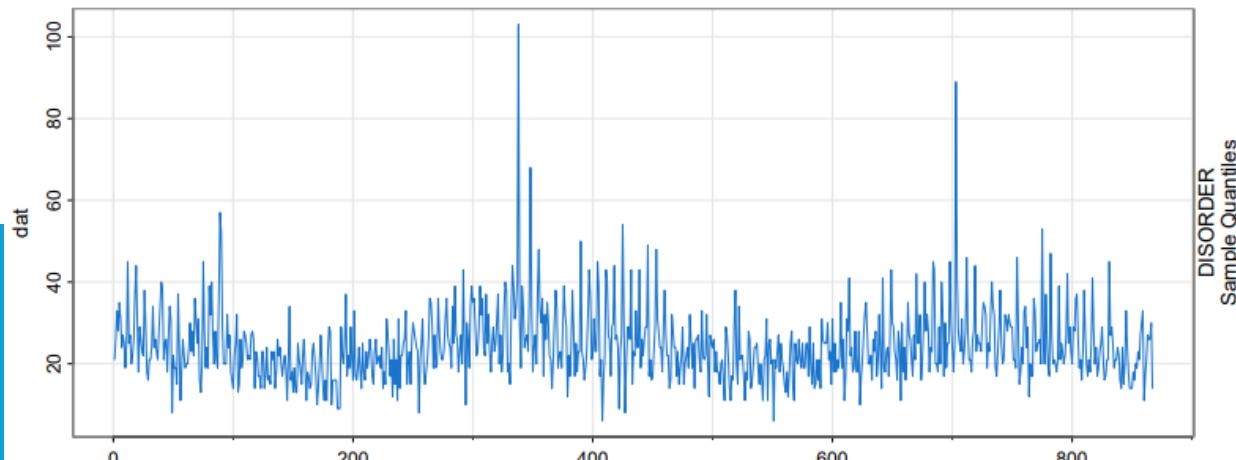


- Analysis Focus
 - Disorder - total recorded 20,981
~10.8% of the total data

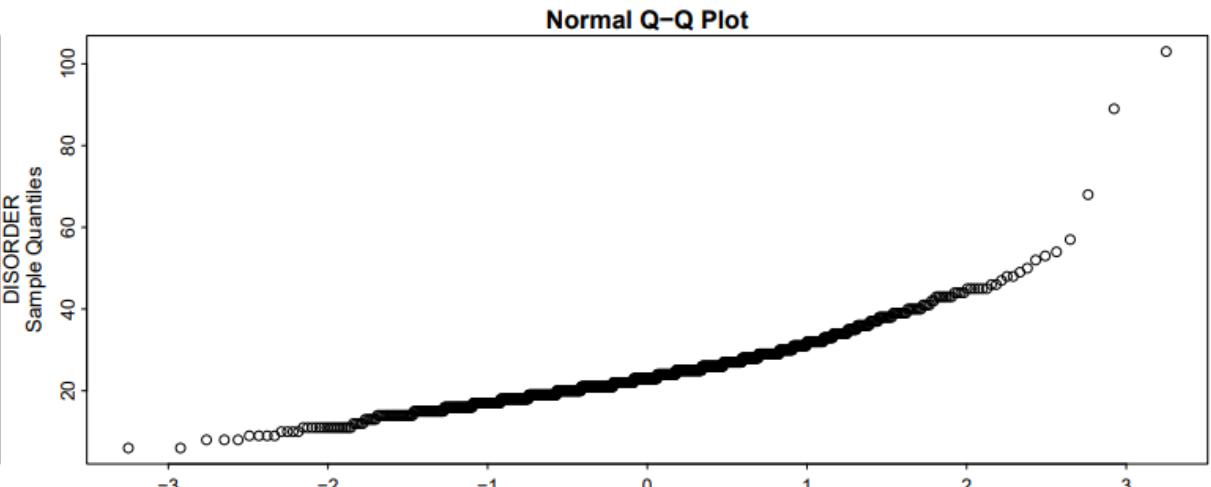
Daily Average: 24

Daily Disorder Incident Counts Diagnostics

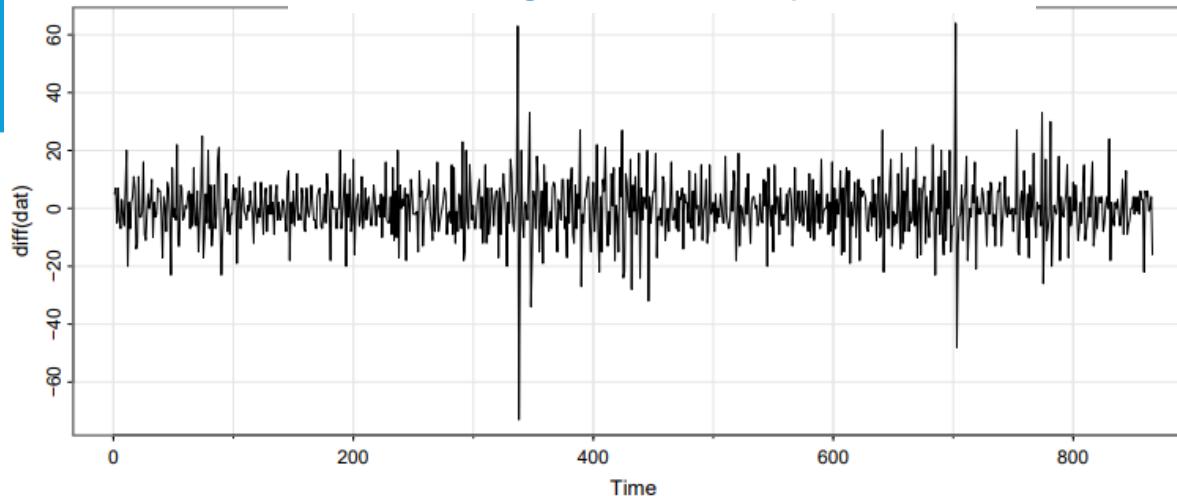
Disorder Daily Counts Time Evolution



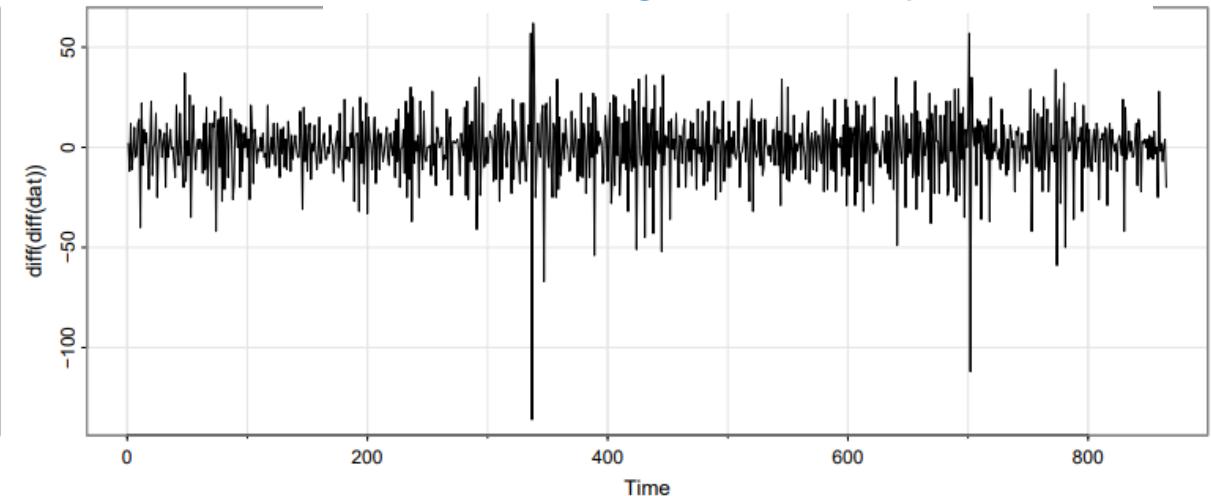
Disorder Daily Counts Distribution



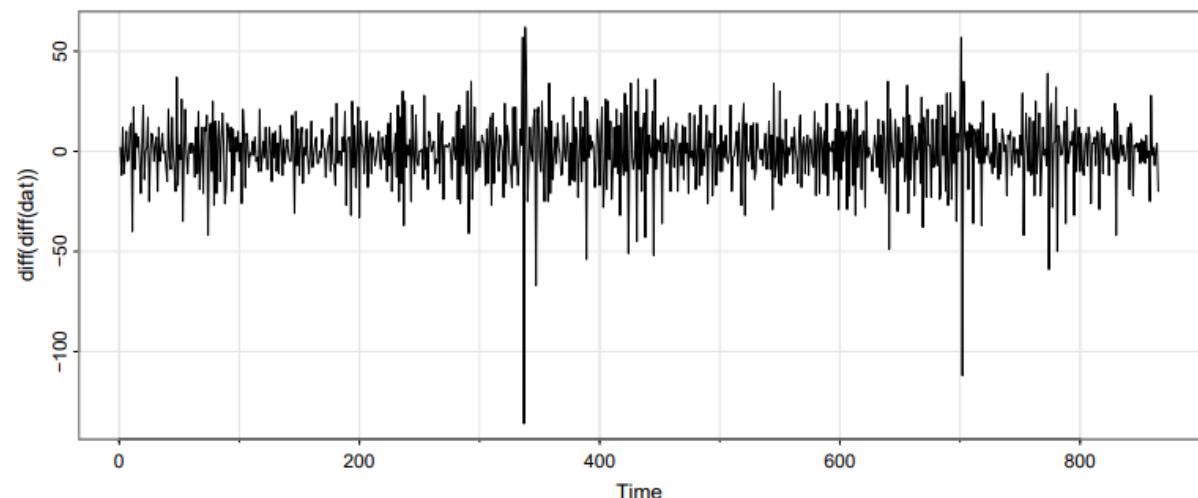
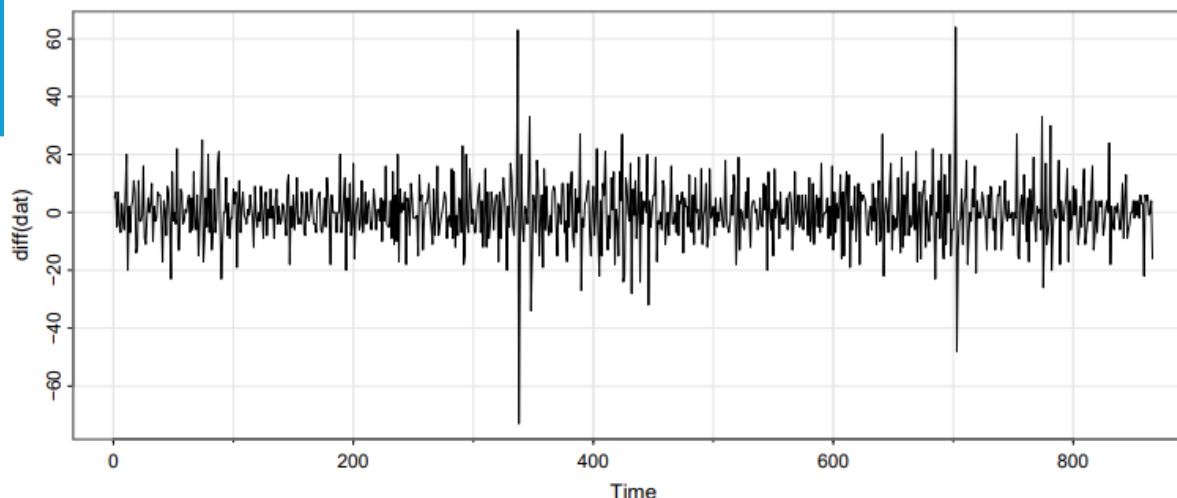
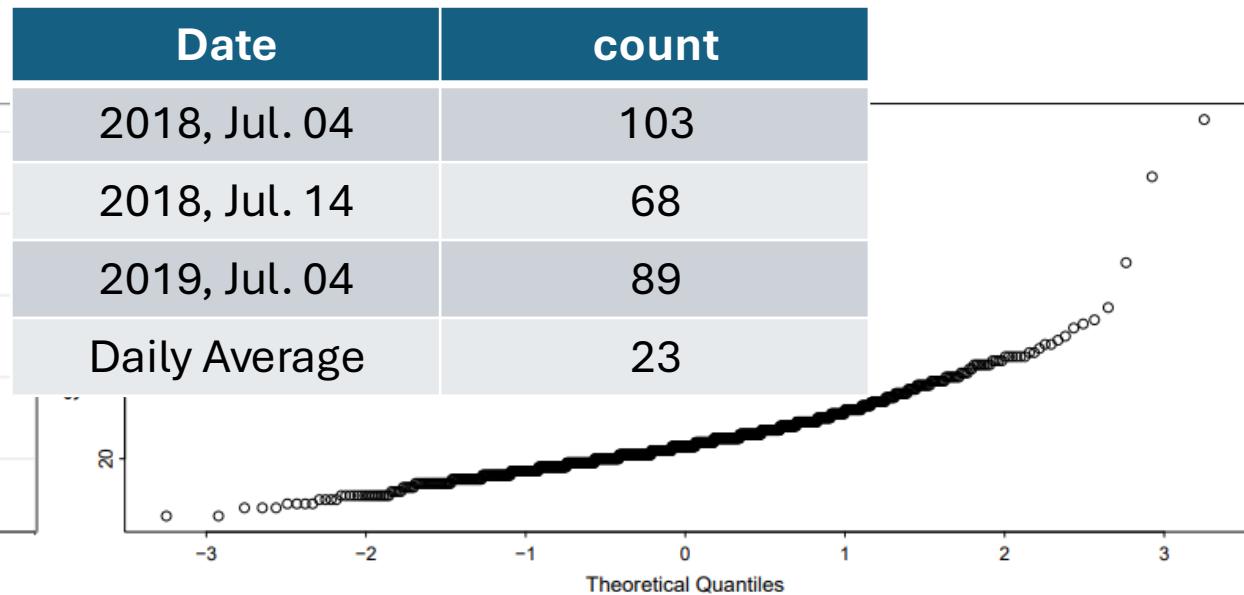
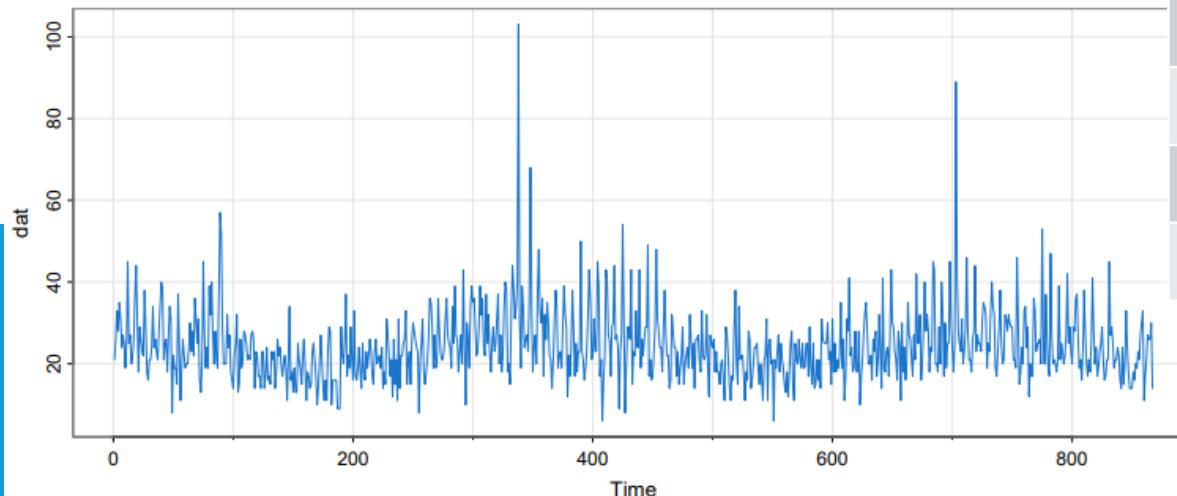
Differencing Disorder Daily Counts



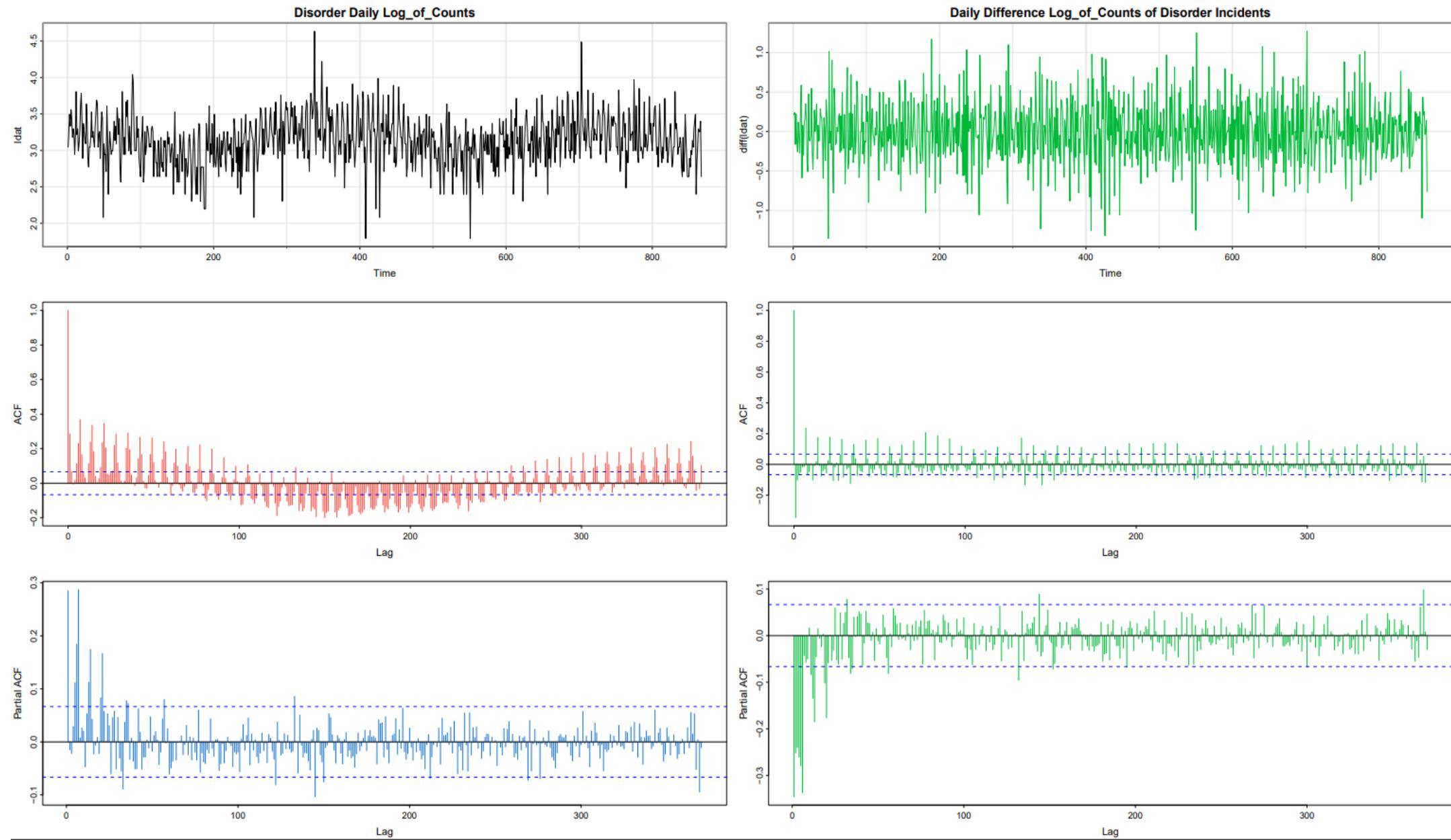
2nd Differencing Disorder Daily Counts



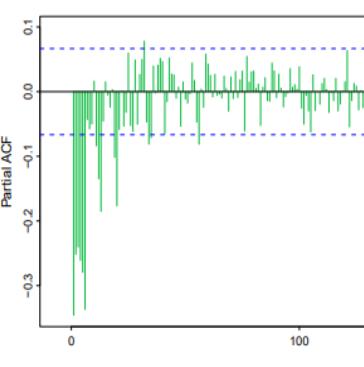
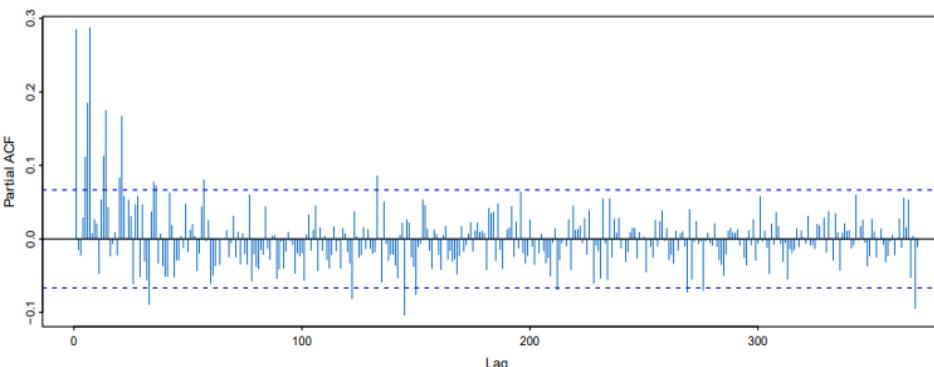
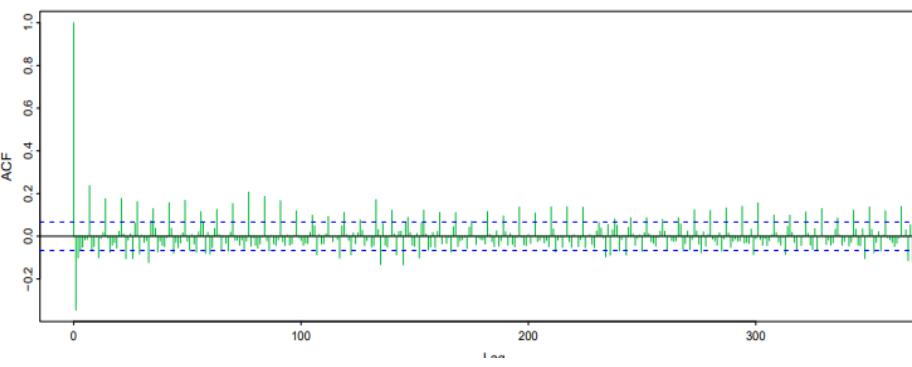
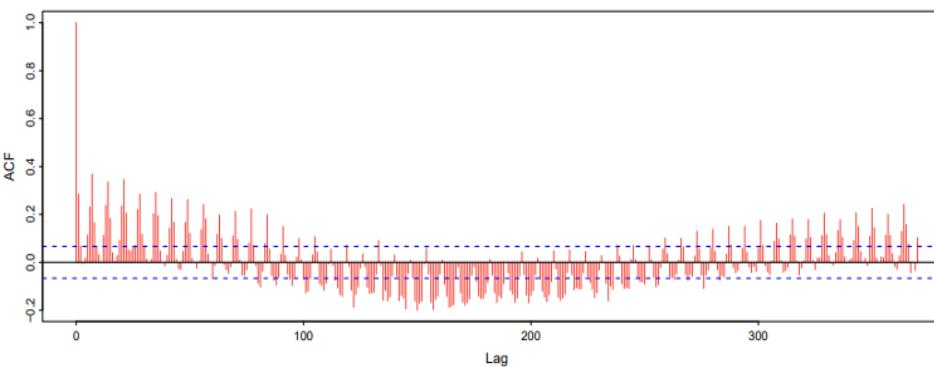
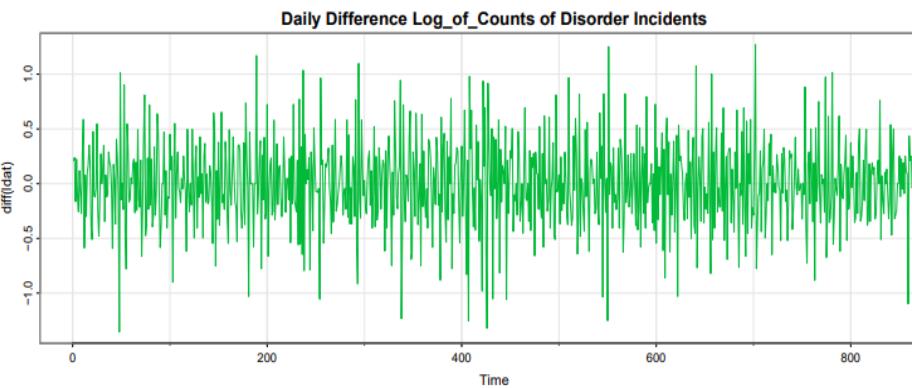
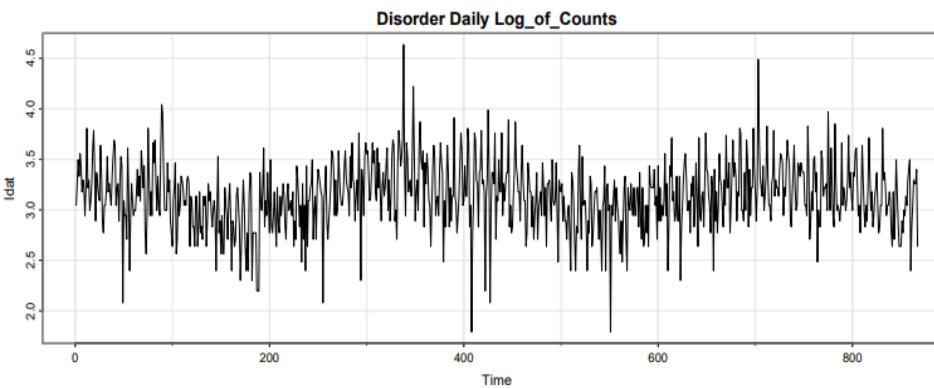
Daily Disorder Incident Counts Diagnostics



Log Transform Daily Disorder Incident Counts



Log Transform Daily Disorder Incident Counts



Coefficients:

	Estimate	SE	t.value	p.value
ar1	0.1323	0.0350	3.7824	2e-04
ma1	-0.9755	0.0079	-122.8428	0e+00
sar1	0.2027	0.0341	5.9437	0e+00
sar2	0.1439	0.0343	4.1910	0e+00
sar3	0.1746	0.0341	5.1210	0e+00

σ^2 estimated as 0.08546242 on 861 degrees of freedom

AIC = 0.3961801 AICc = 0.3962606 BIC = 0.4291862

Disorder Day Log(Count) Model

SARIMA (**log(counts)**)

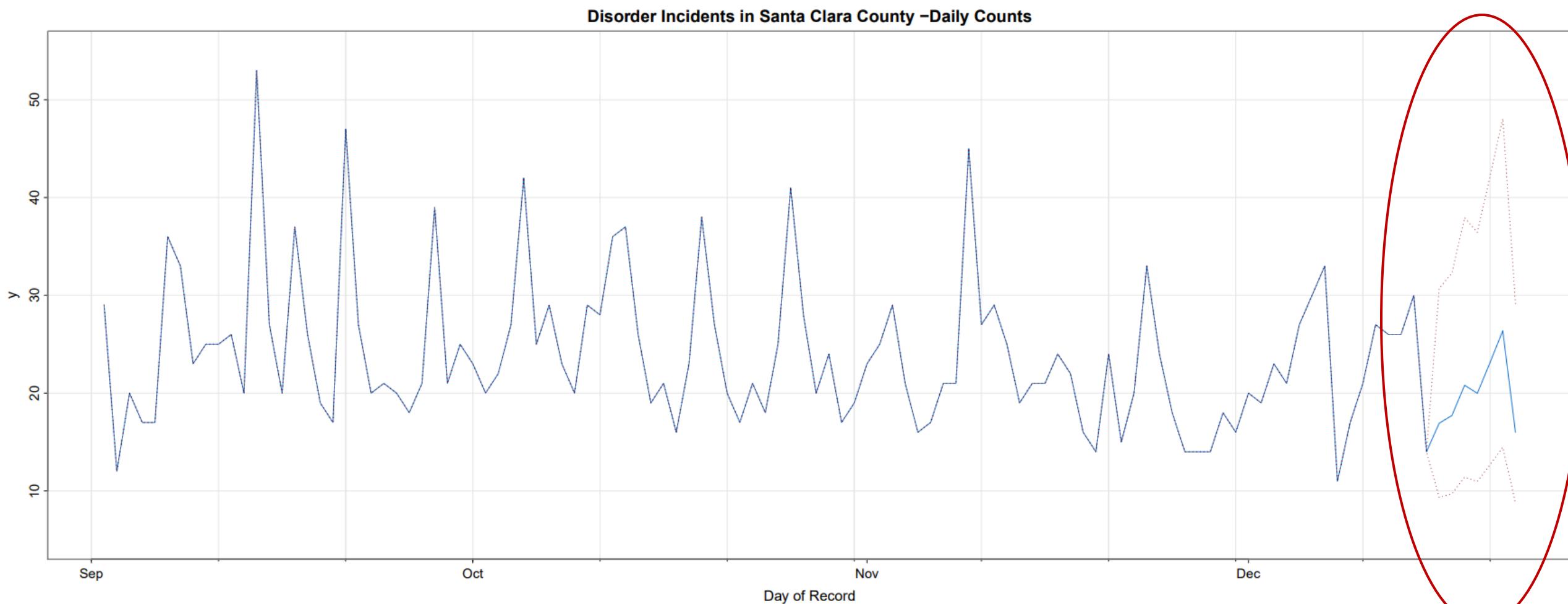
p=1, d=1, q=1

P=3, D=1, Q=0

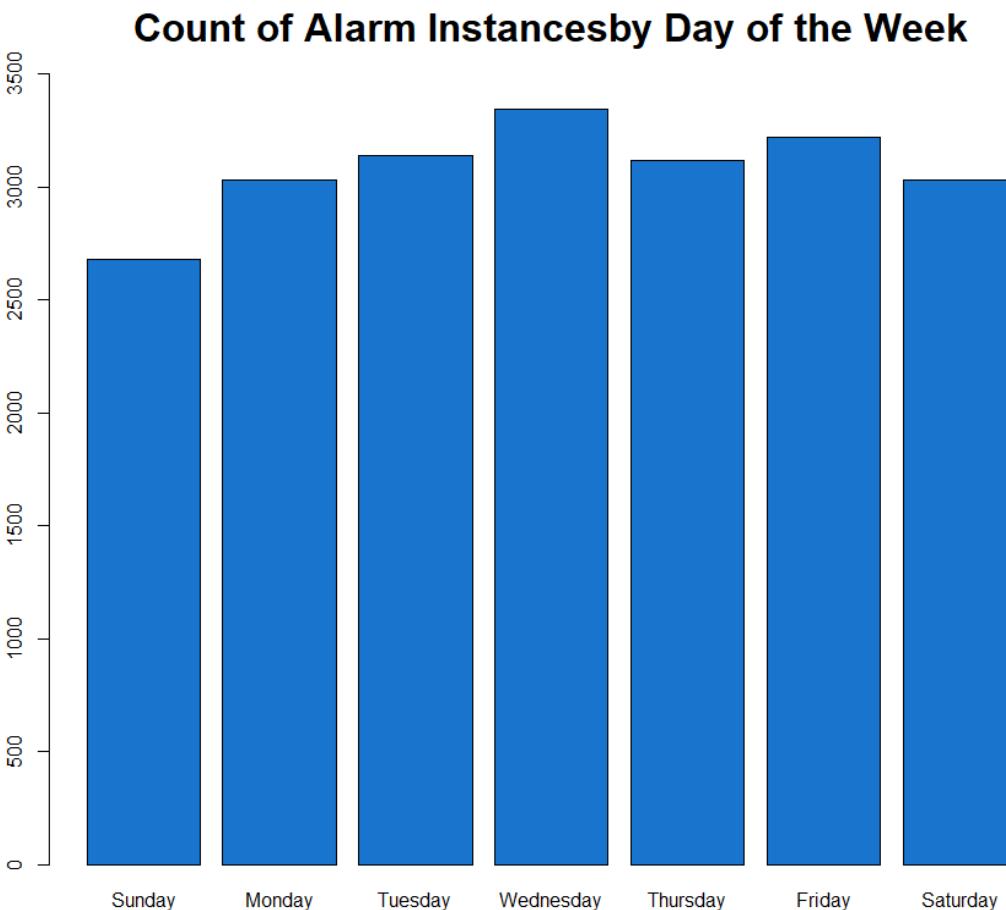
7-day seasonality

Events of 7, 8, 9, and 10 ago have stronger influence than the events of yesterday

Disorder Incidents Seven-Day Prediction for Dec. 2019



Daily Trends: Alarms

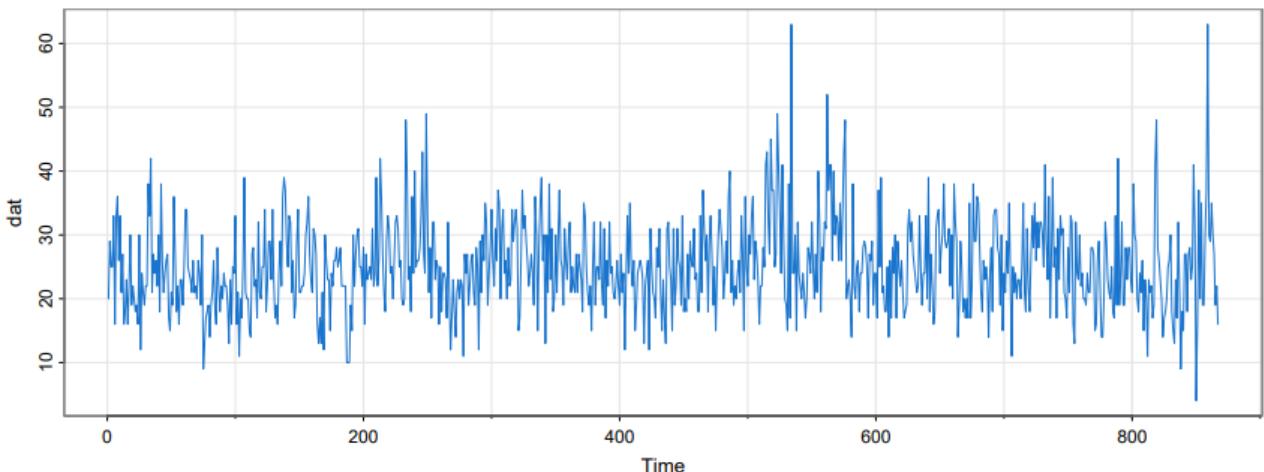


- **Analysis Focus**
on preventive policing incidents, each with more than 10 counts daily
 - Disorder - total recorded 21, 563
~11.1% of the total data

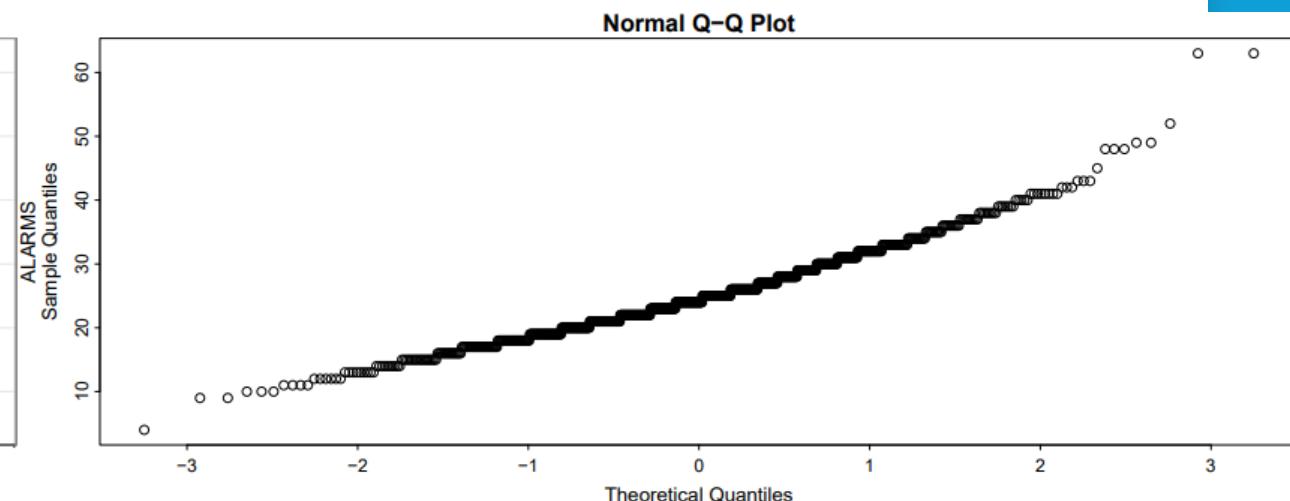
Daily Average: 25

Daily Alarm Incident Counts Diagnostics

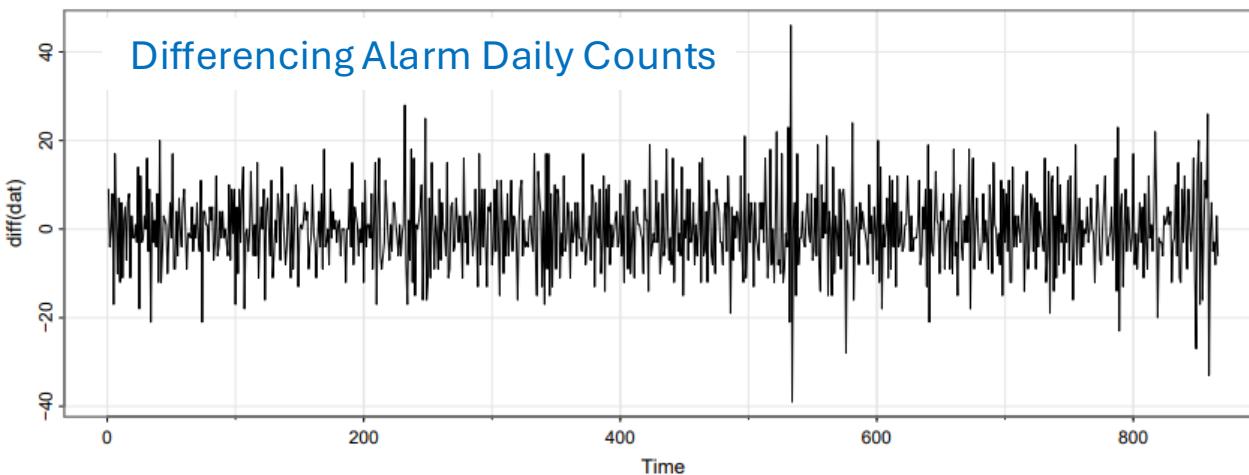
Alarm Daily Counts Time Evolution



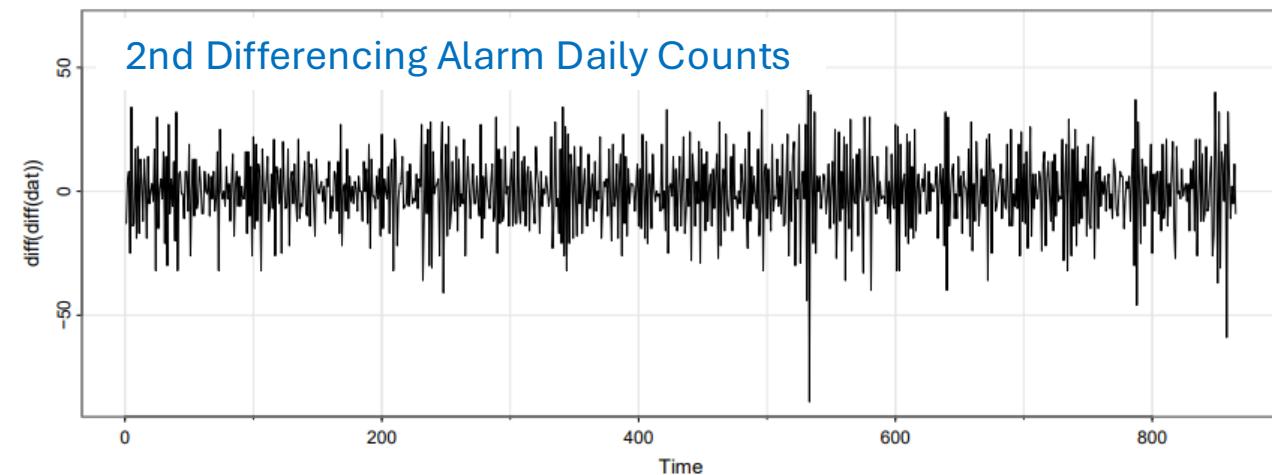
Alarm Daily Counts Distribution



Differencing Alarm Daily Counts

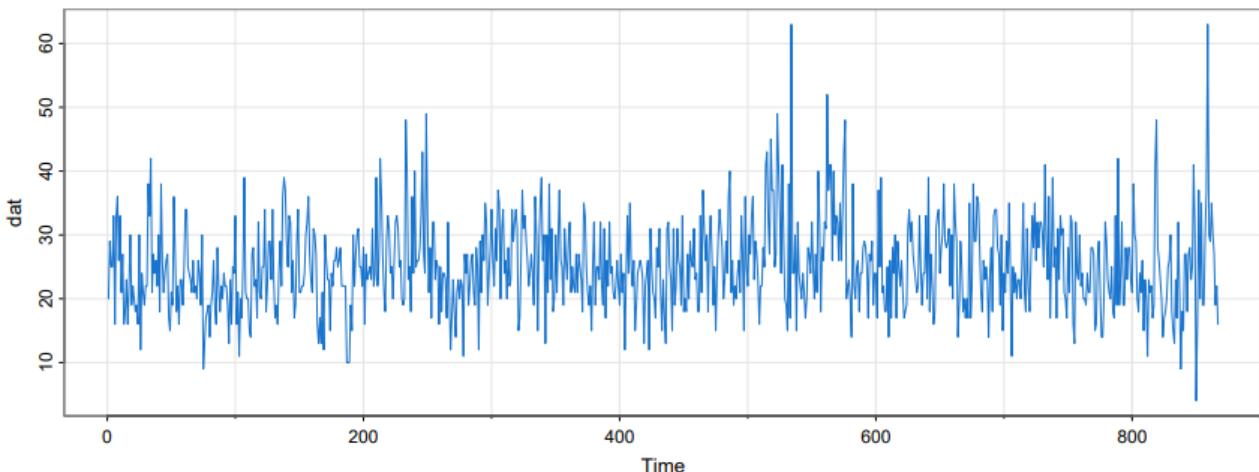


2nd Differencing Alarm Daily Counts

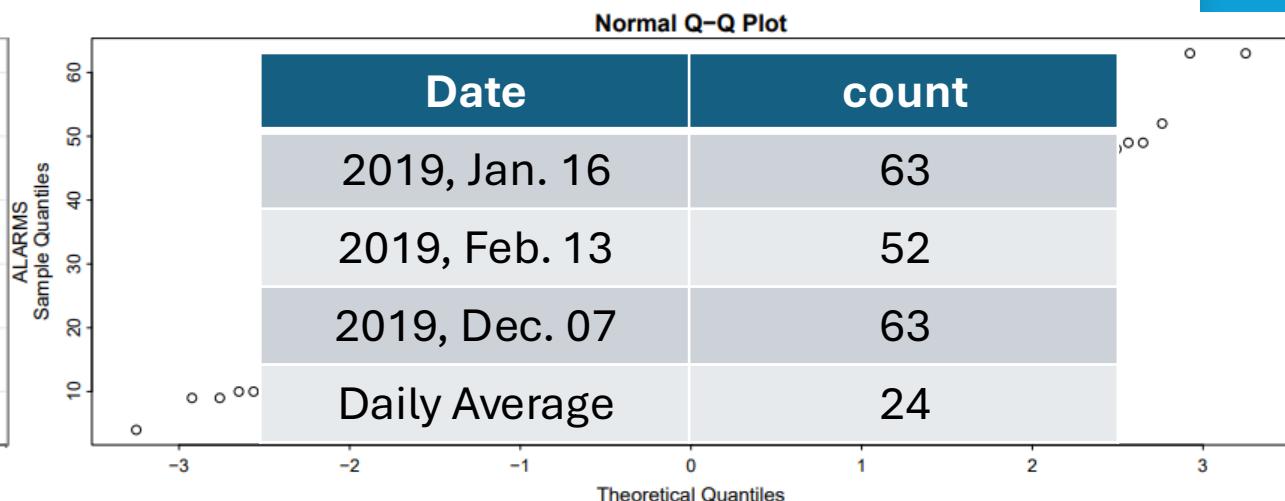


Daily Alarm Incident Counts Diagnostics

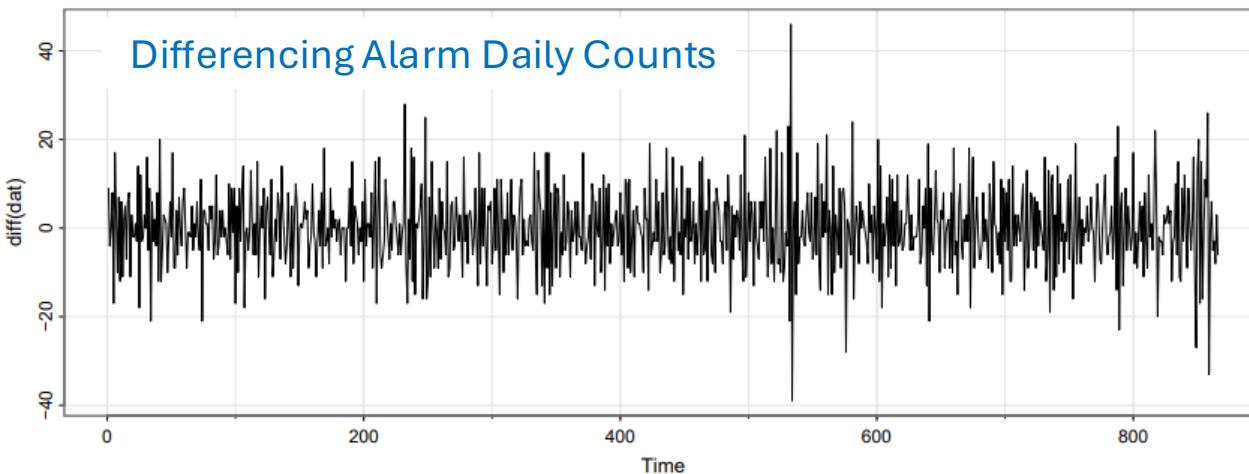
Alarm Daily Counts Time Evolution



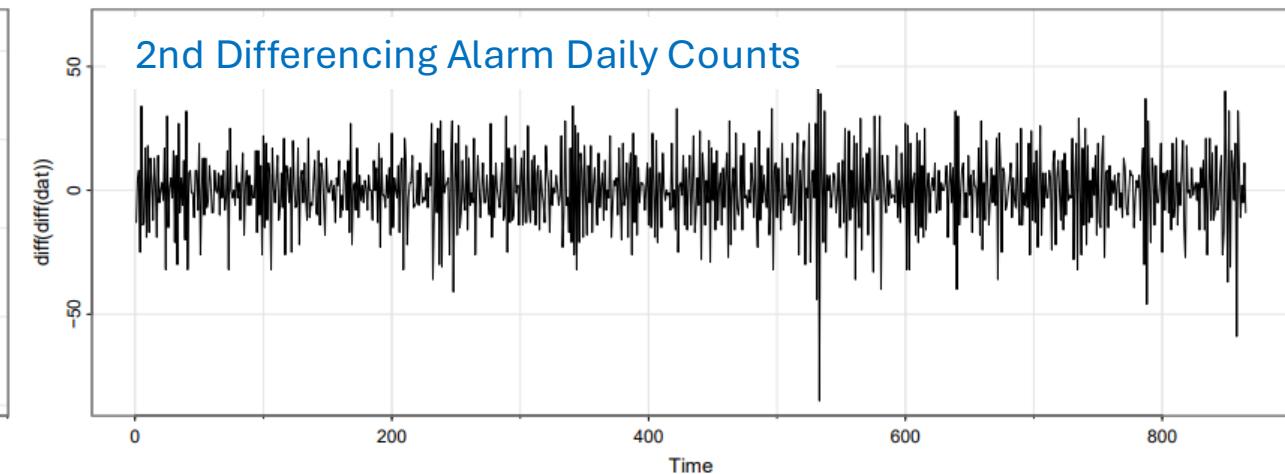
Alarm Daily Counts Distribution



Differencing Alarm Daily Counts

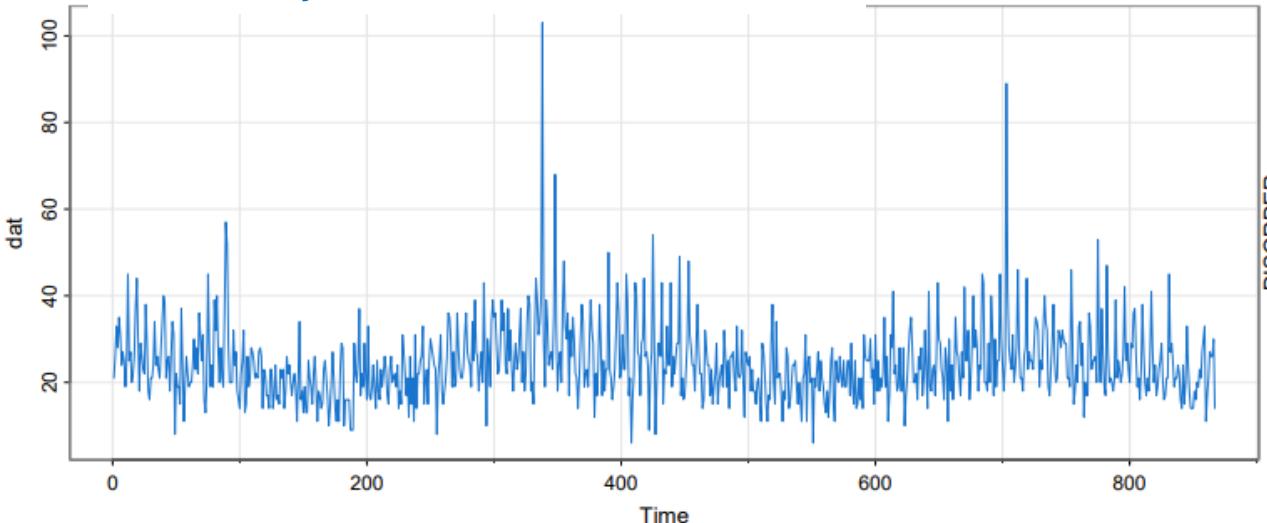


2nd Differencing Alarm Daily Counts

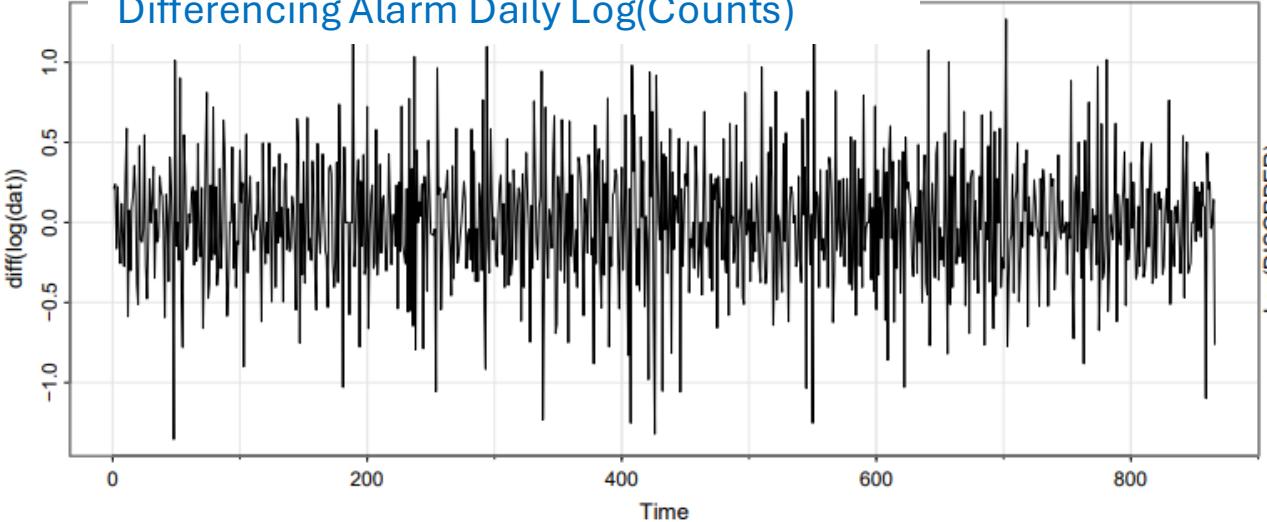


Daily Alarm Incident Log(Counts) Diagnostics

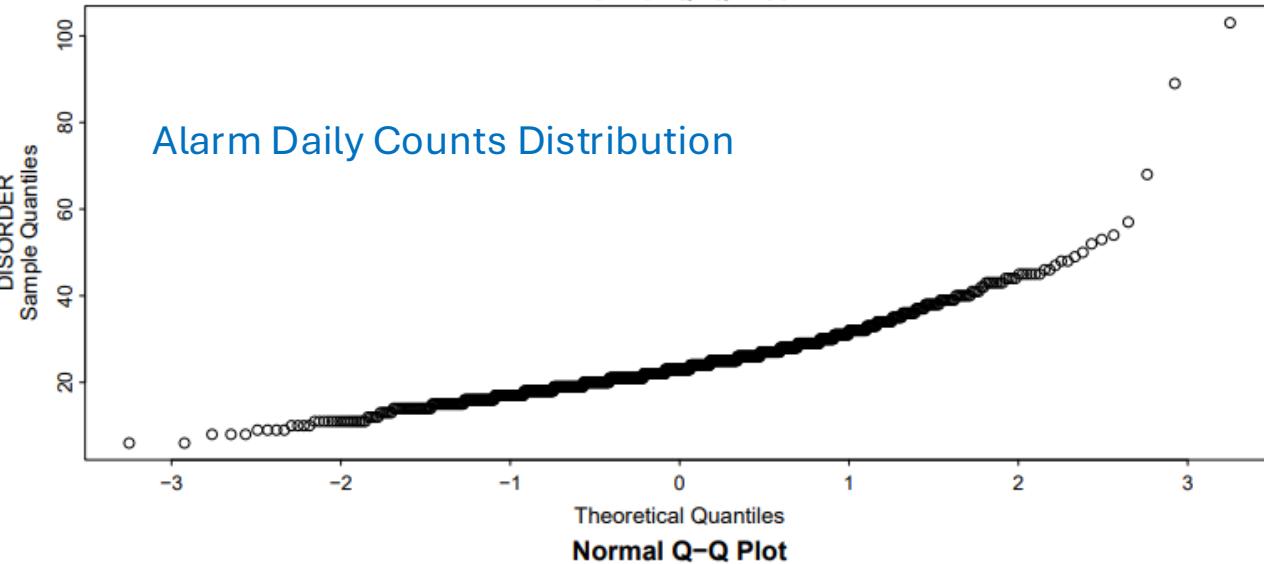
Alarm Daily Counts Time Evolution



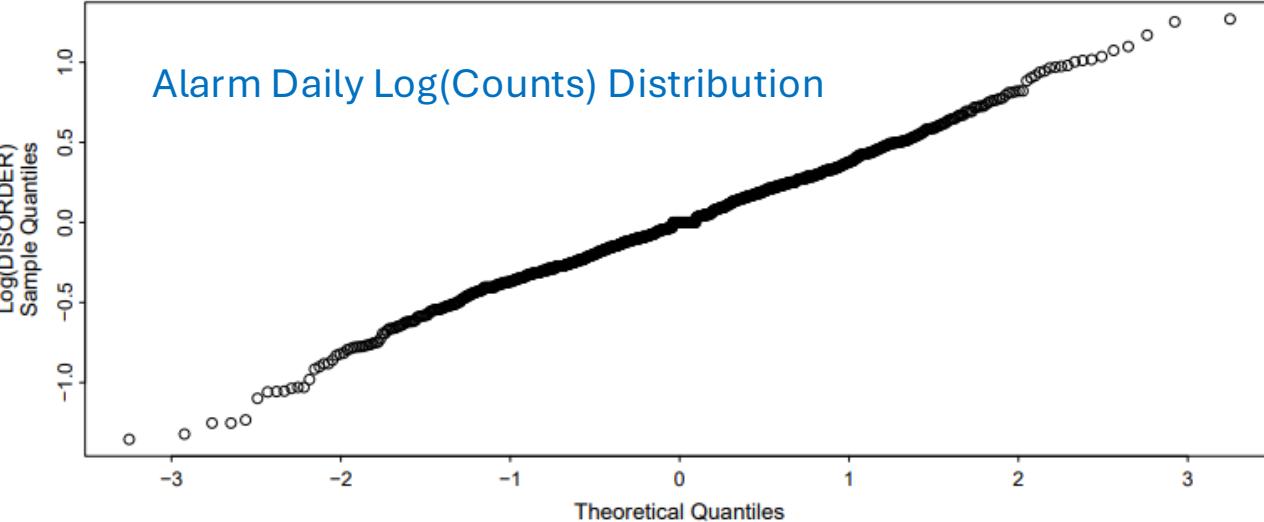
Differencing Alarm Daily Log(Counts)



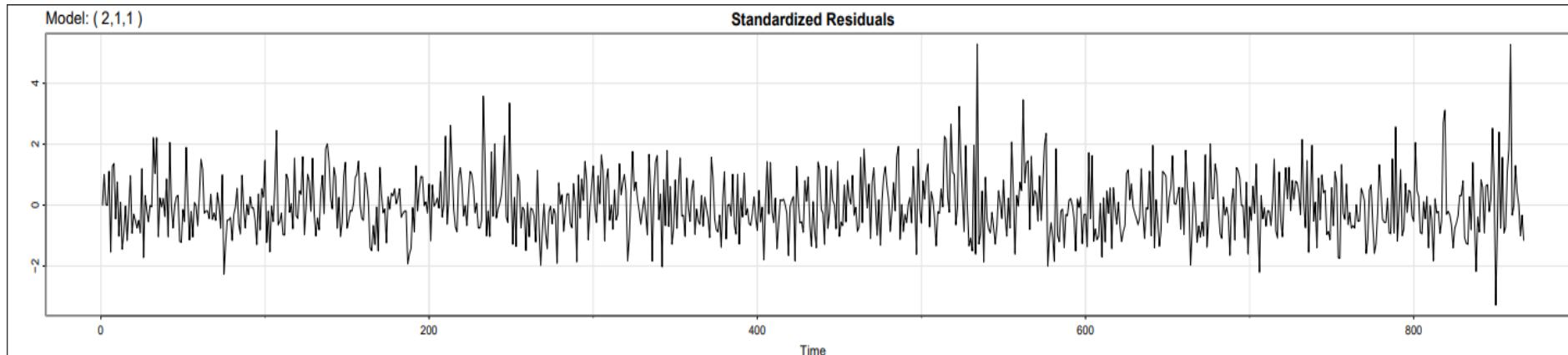
Normal Q-Q Plot



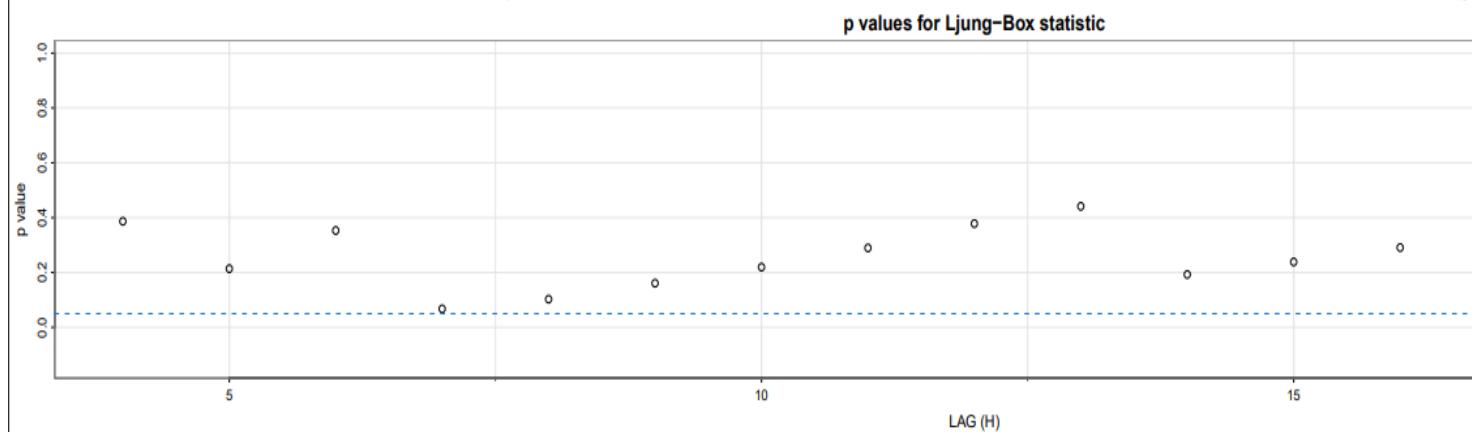
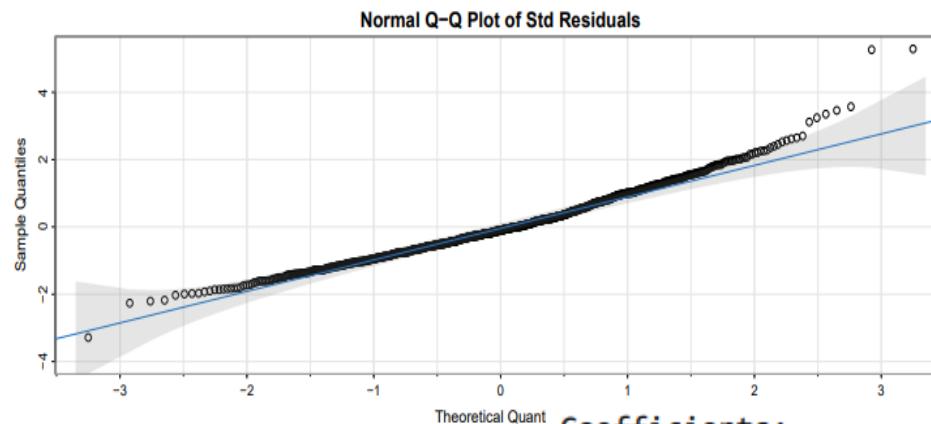
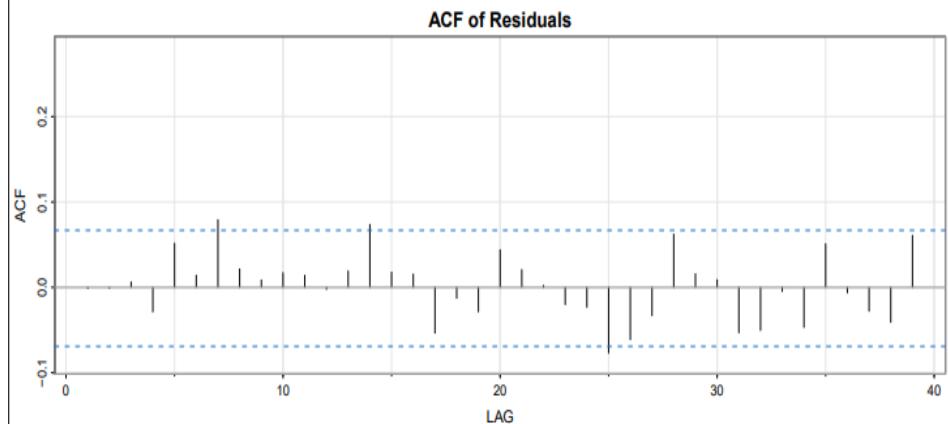
Normal Q-Q Plot



Daily Alarm Incident Model & Goodness of Fit



Alarm Day Log(Count) Model



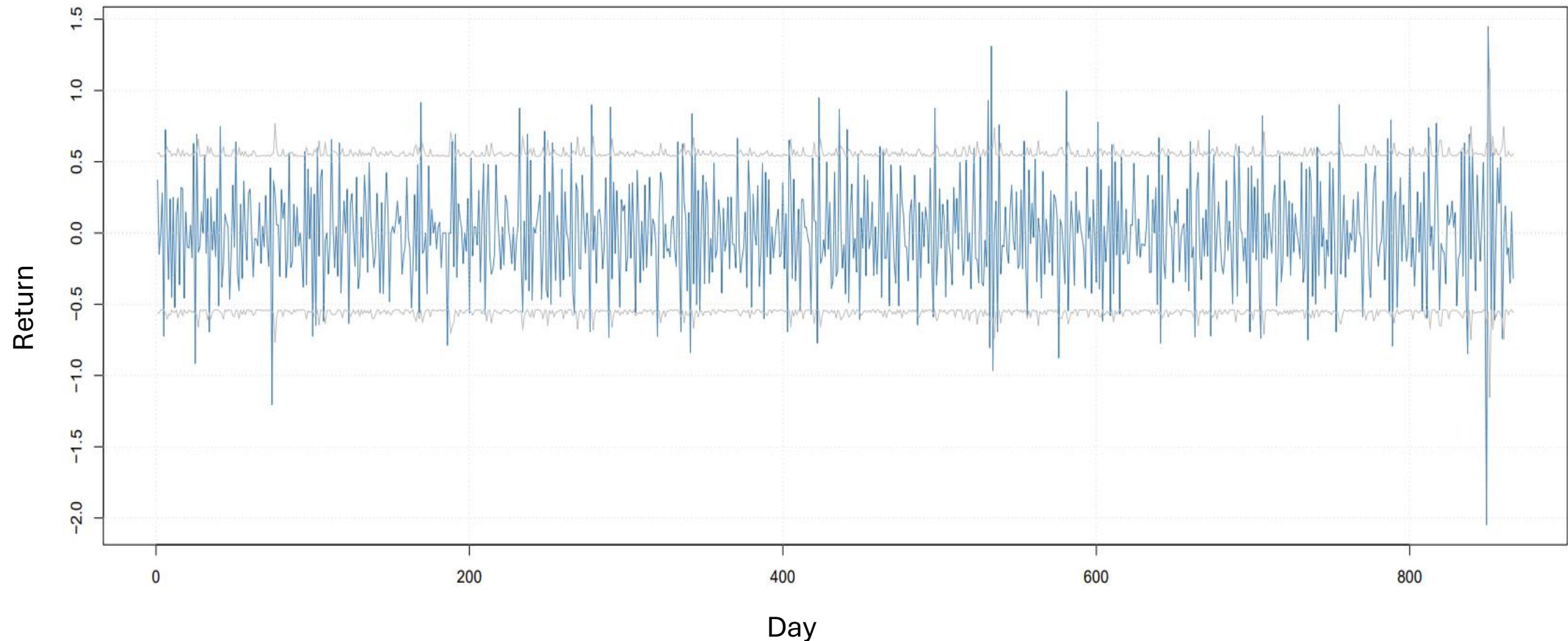
	Coefficients:	Estimate	SE	t.value	p.value
ar1	0.1664	0.0351	4.7420	0.0000	
ar2	0.0932	0.0351	2.6574	0.0080	
ma1	-0.9877	0.0094	-104.8394	0.0000	
sar1	0.0849	0.0349	2.4354	0.0151	

σ^2 estimated as 47.16347 on 862 degrees of freedom

AIC = 6.706583 AICc = 6.706637 BIC = 6.734088

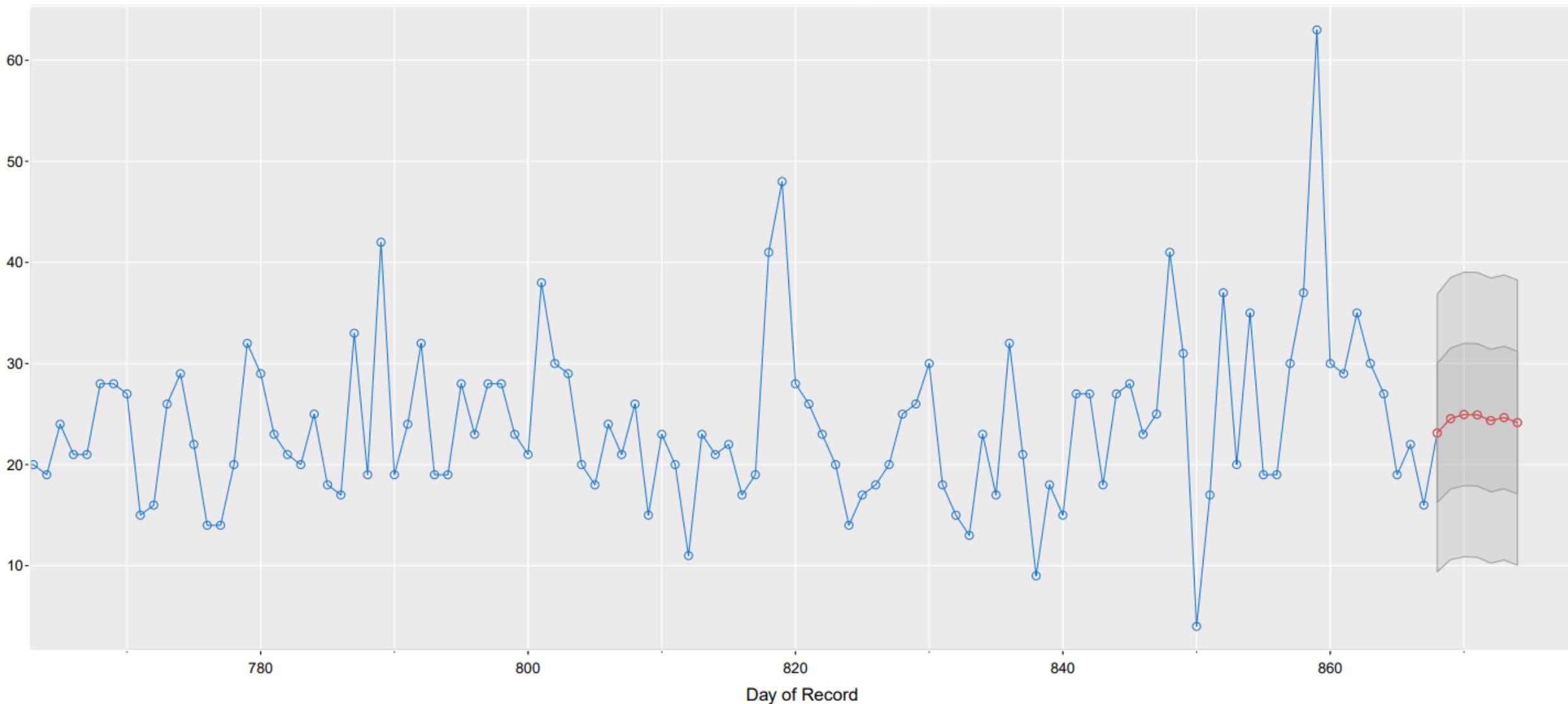
Daily Alarms Conditional Variance through GARCH

Series with 2 Conditional SD Superimposed

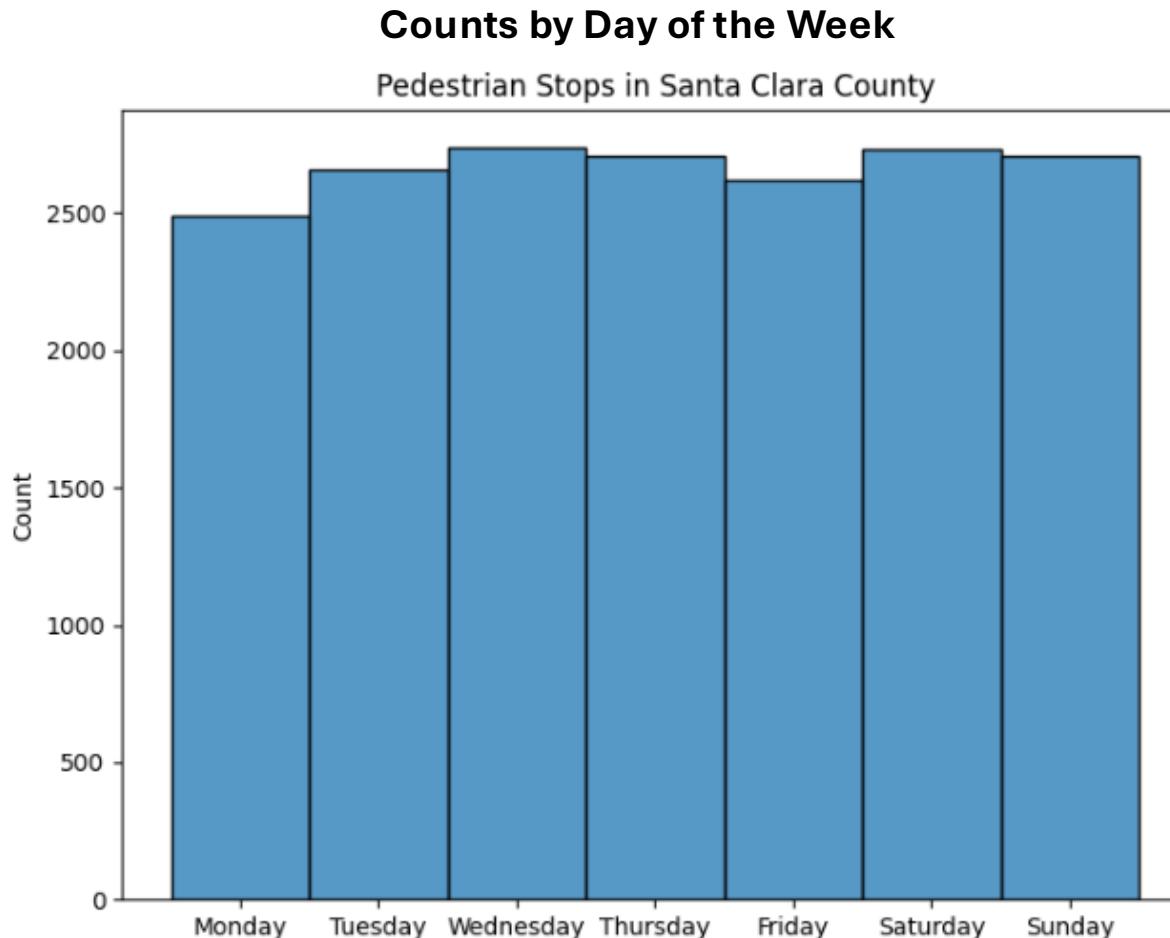


Alarm Incidents Seven-Day Prediction for Dec. 2019

Alarms in Santa Clara County – Daily Counts



Daily Trends: Pedestrian Stops

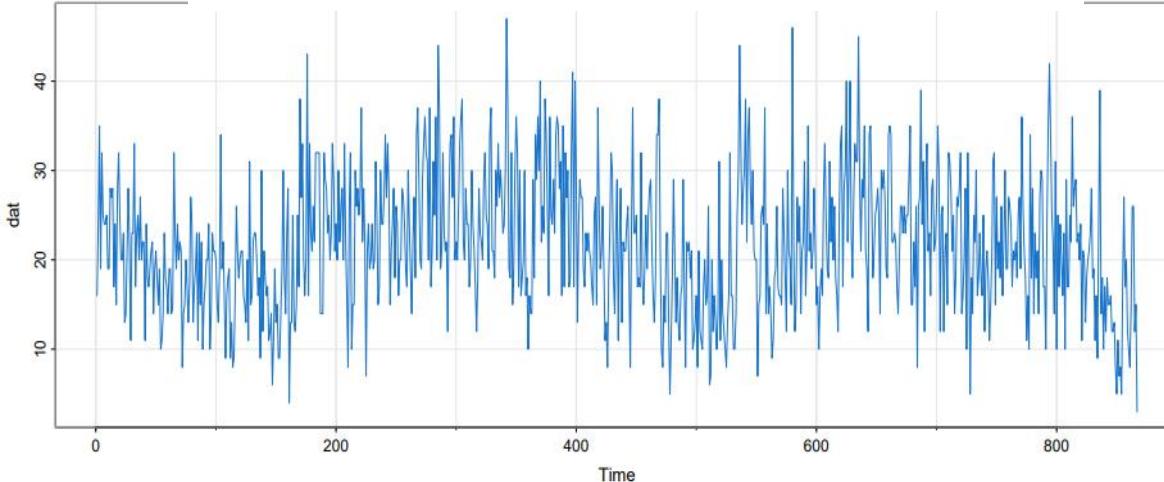


- **Analysis Focus**
on preventive policing incidents, each
with more than 10 counts daily
- Pedestrian Stops - total recorded
18,652
~ 9.6% of the total data

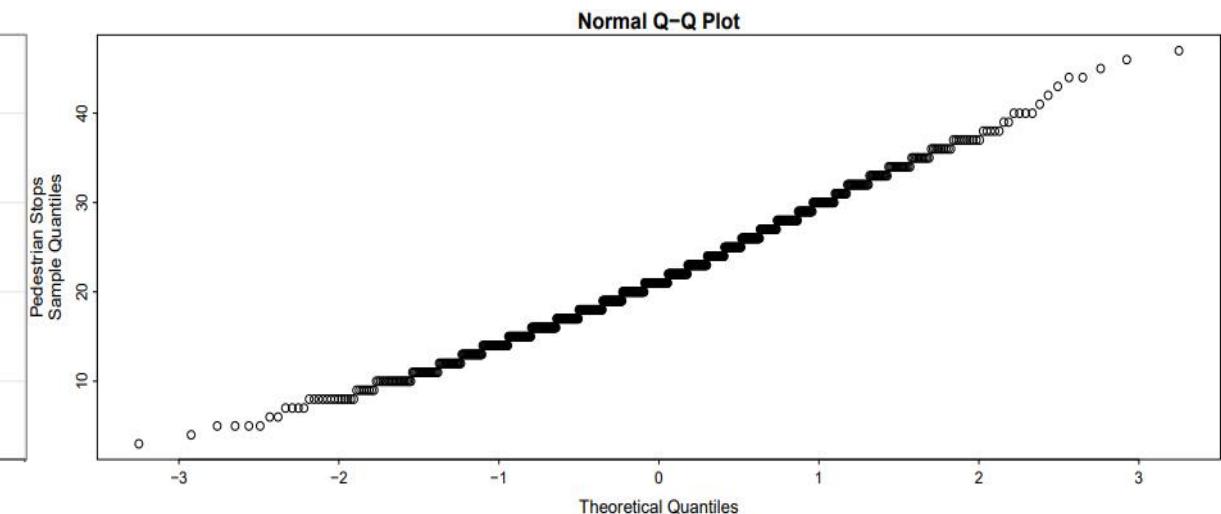
Daily Average: 22

Daily Pedestrian Stop Counts Diagnostics

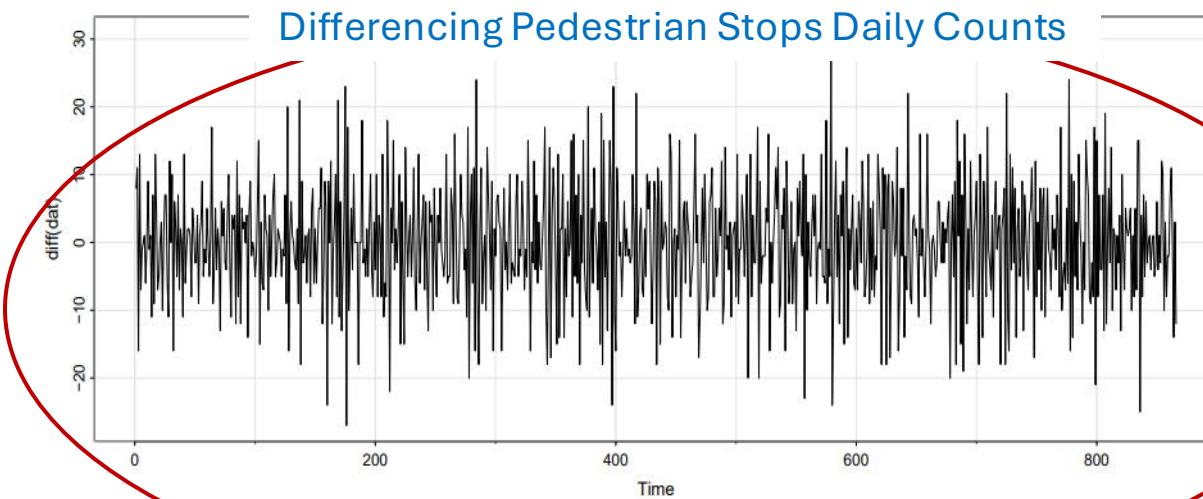
Pedestrian Stop Daily Counts Time Evolution



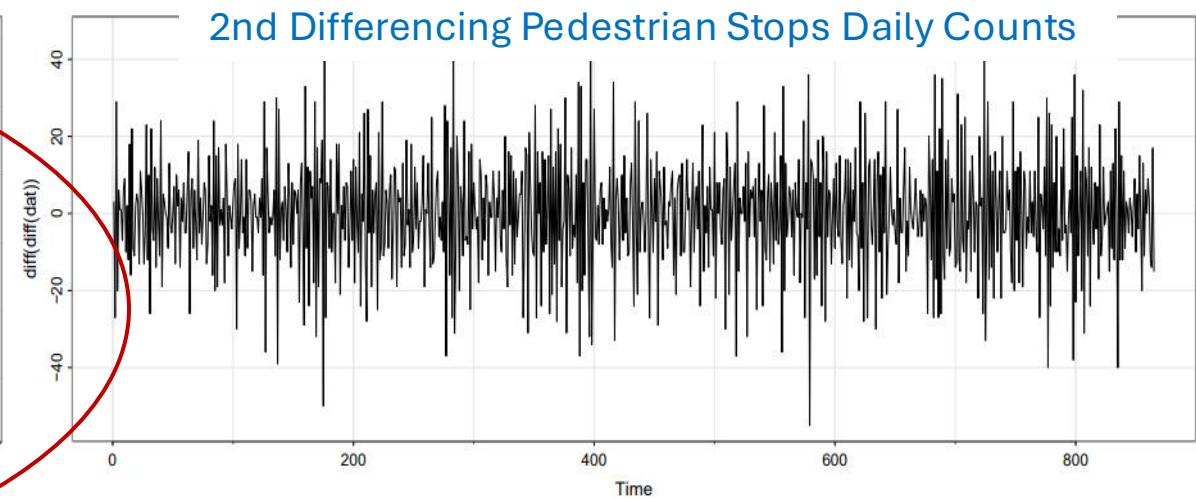
Pedestrian Stops Daily Count Distribution



Differencing Pedestrian Stops Daily Counts

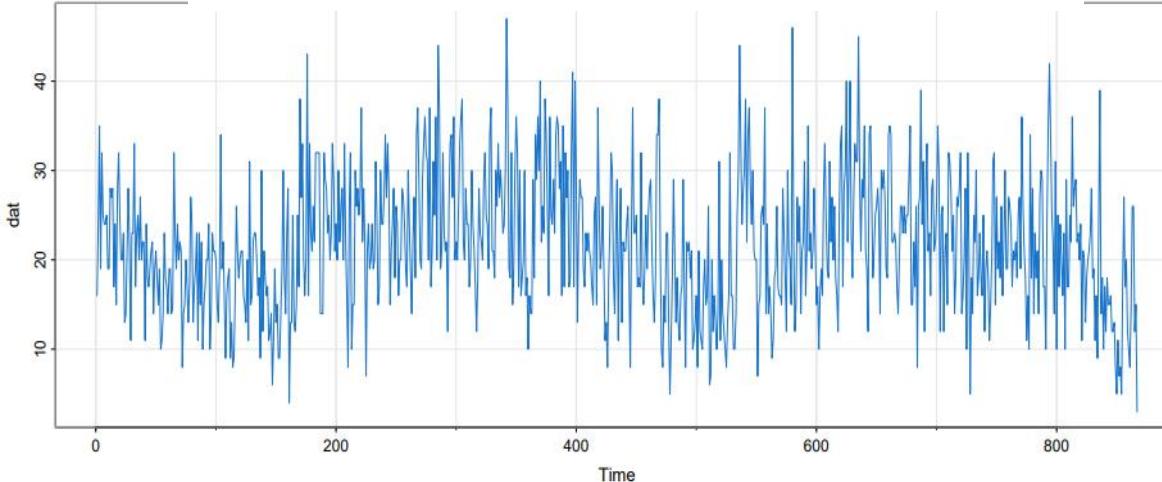


2nd Differencing Pedestrian Stops Daily Counts

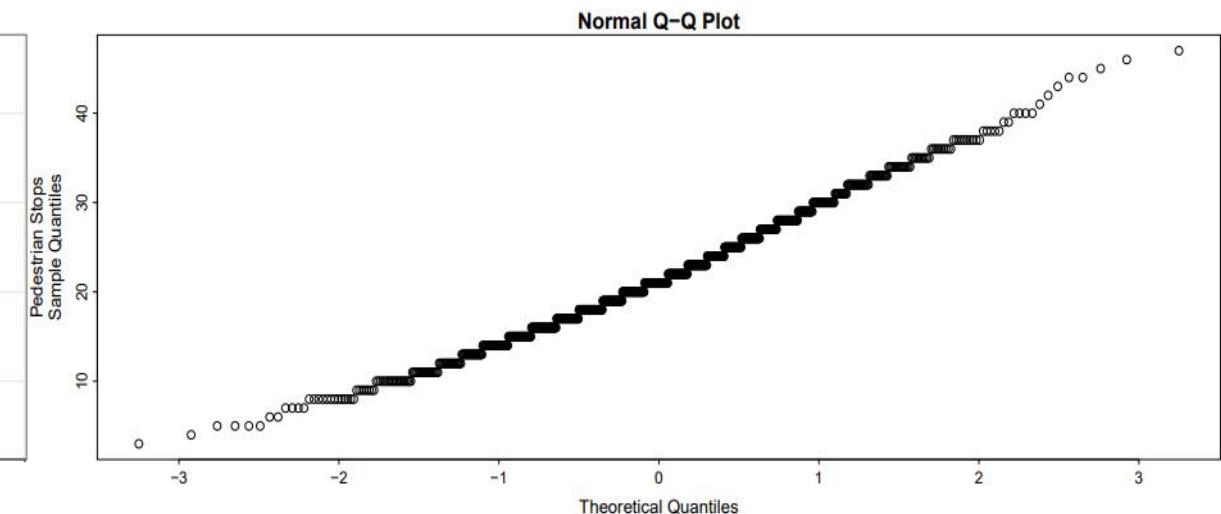


Daily Pedestrian Stop Counts Diagnostics

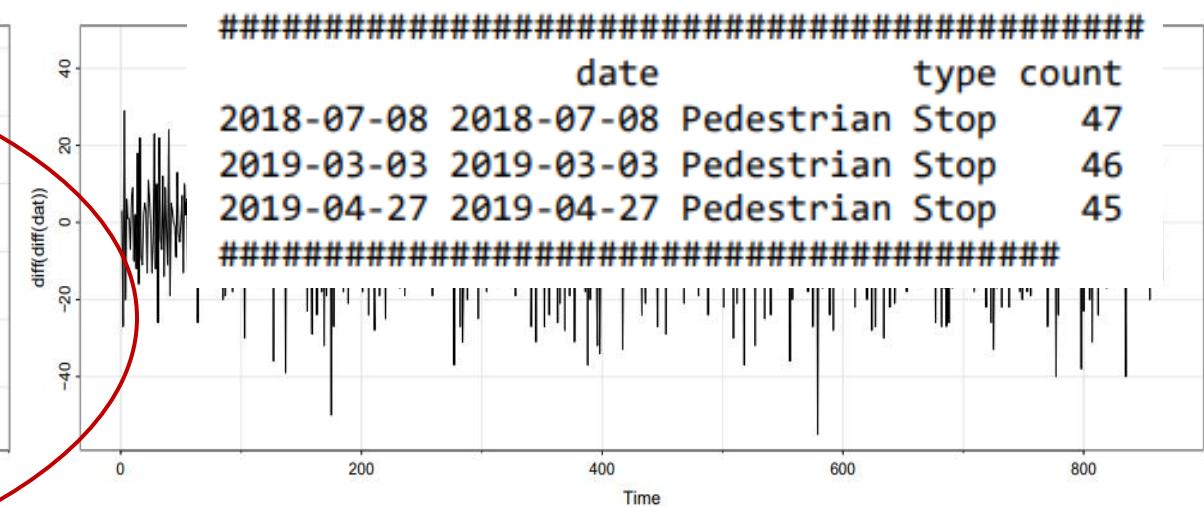
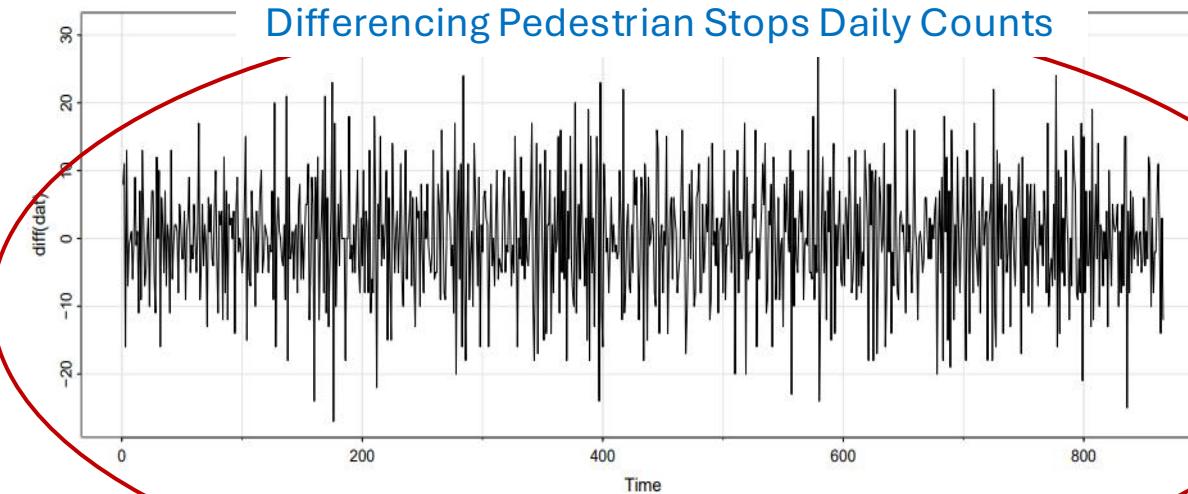
Pedestrian Stop Daily Counts Time Evolution



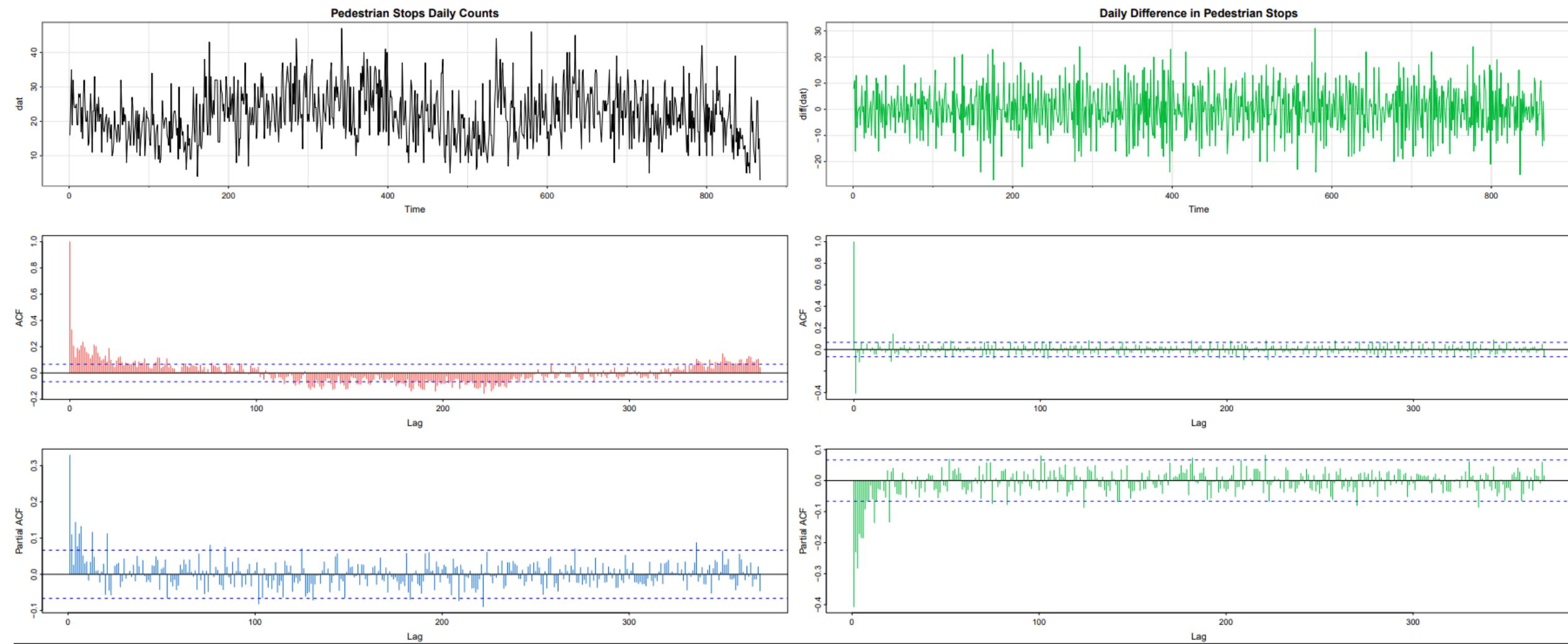
Pedestrian Stops Daily Count Distribution



Differencing Pedestrian Stops Daily Counts



Daily Pedestrian Stops Diagnostics



Daily Pedestrian Stops Model Goodness of Fit

Pedestrian Stops Day Count Model

SARIMA (**log(counts)**)

p=1, d=1, q=2

P=0, D=0, Q=1

7-day seasonality

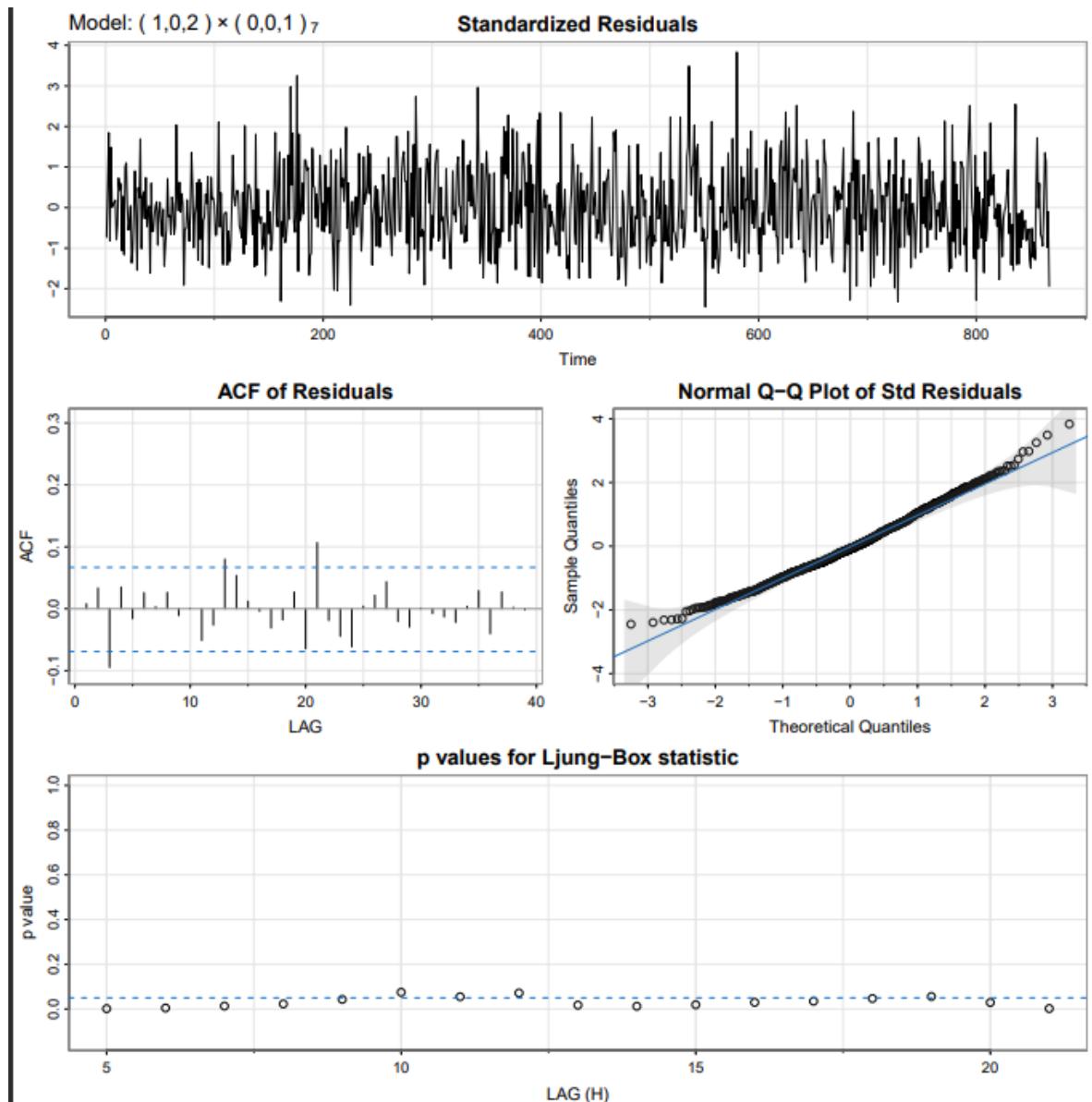
```
sarima(dat, 1, 0, 2, 0, 0, 1, 7)
```

Coefficients:

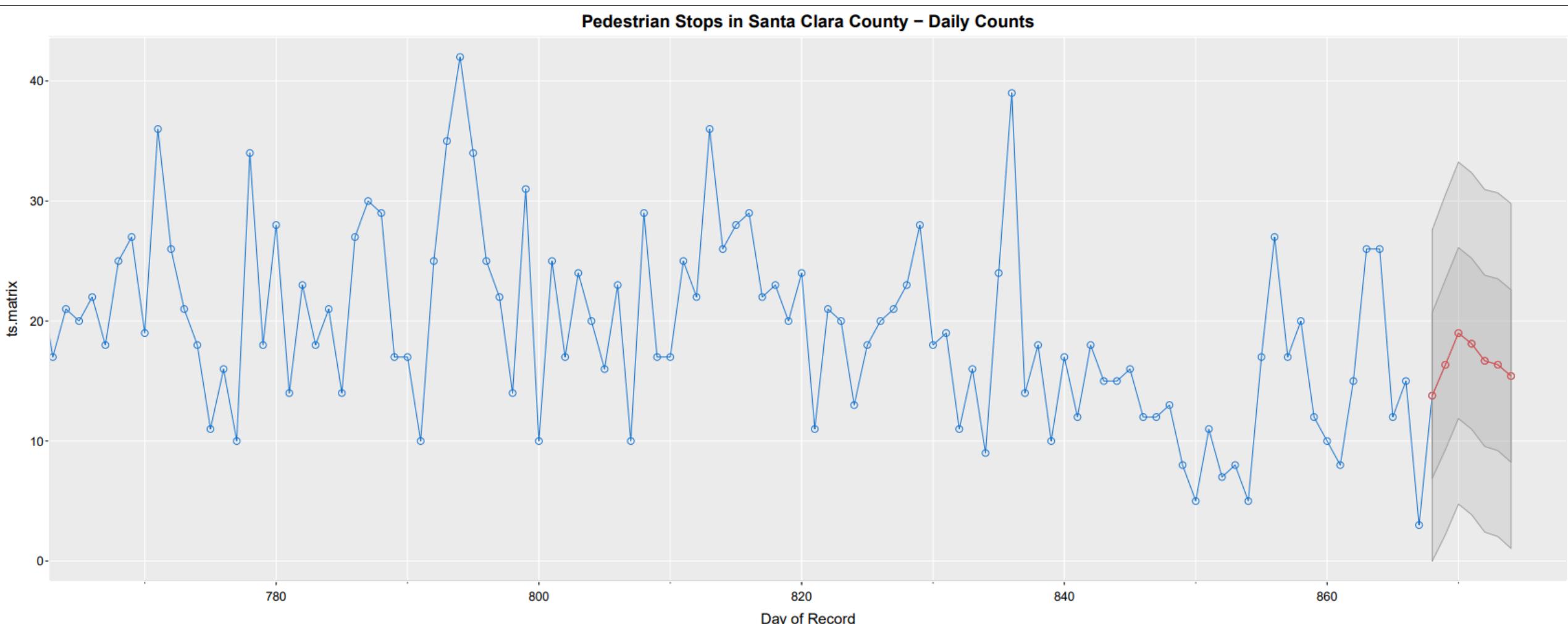
	Estimate	SE	t.value	p.value
ar1	0.9773	0.0126	77.5843	0.0000
ma1	-0.7540	0.0350	-21.5217	0.0000
ma2	-0.1421	0.0326	-4.3608	0.0000
sma1	0.0671	0.0337	1.9928	0.0466
xmean	21.5662	1.1182	19.2857	0.0000

σ^2 estimated as 48.42082 on 862 degrees of freedom

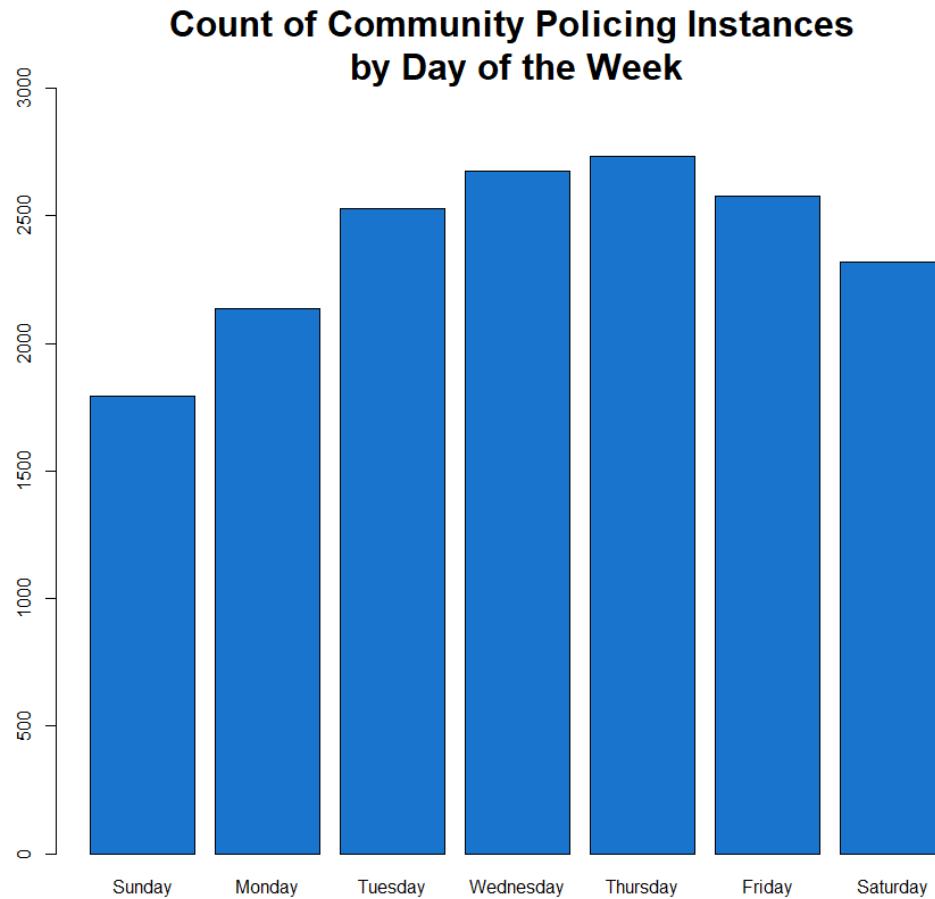
AIC = 6.732315 AICc = 6.732395 BIC = 6.765291



Pedestrian Stops Seven-Day Prediction for Dec. 2019



Daily Trends: Community Policing



- Analysis Focus

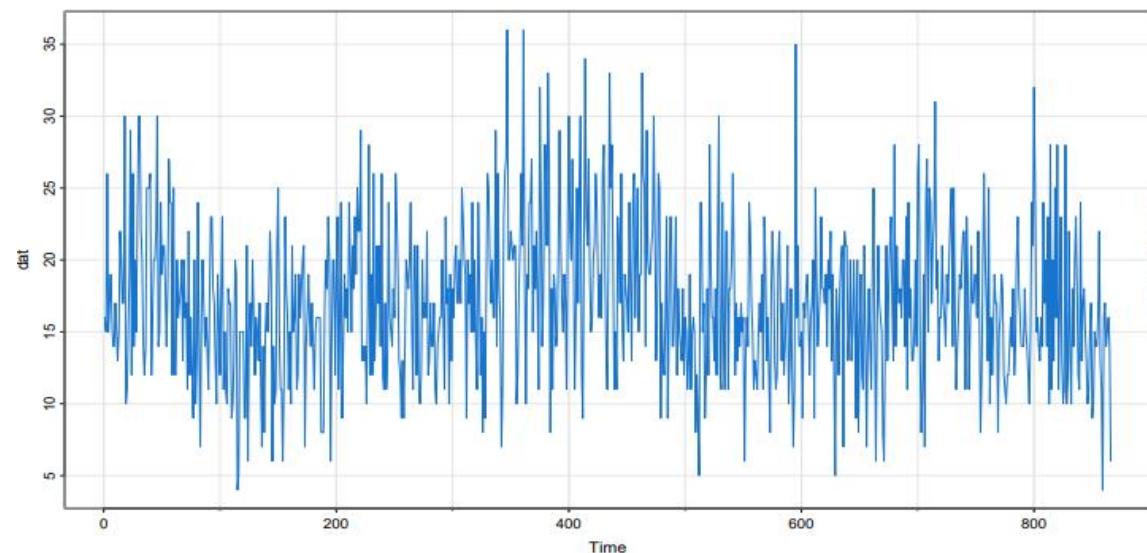
on preventive policing incidents, each with more than 10 counts daily

- Disorder - total recorded 14,635
~7.5% of the total data

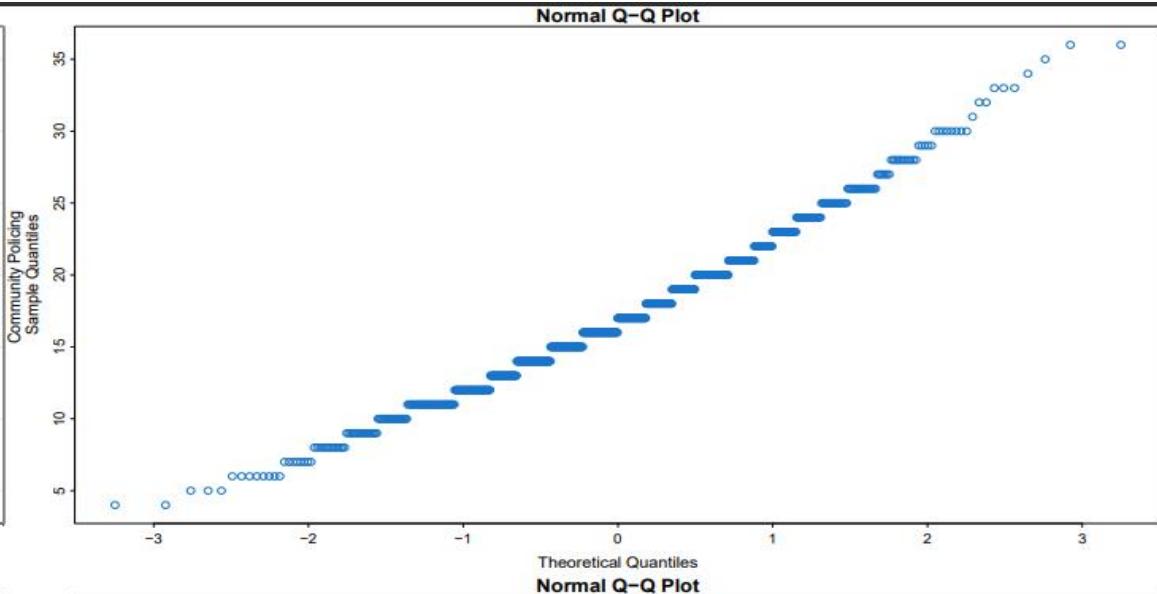
Daily Average: 17

Daily Community Policing Incident Counts Diagnostics

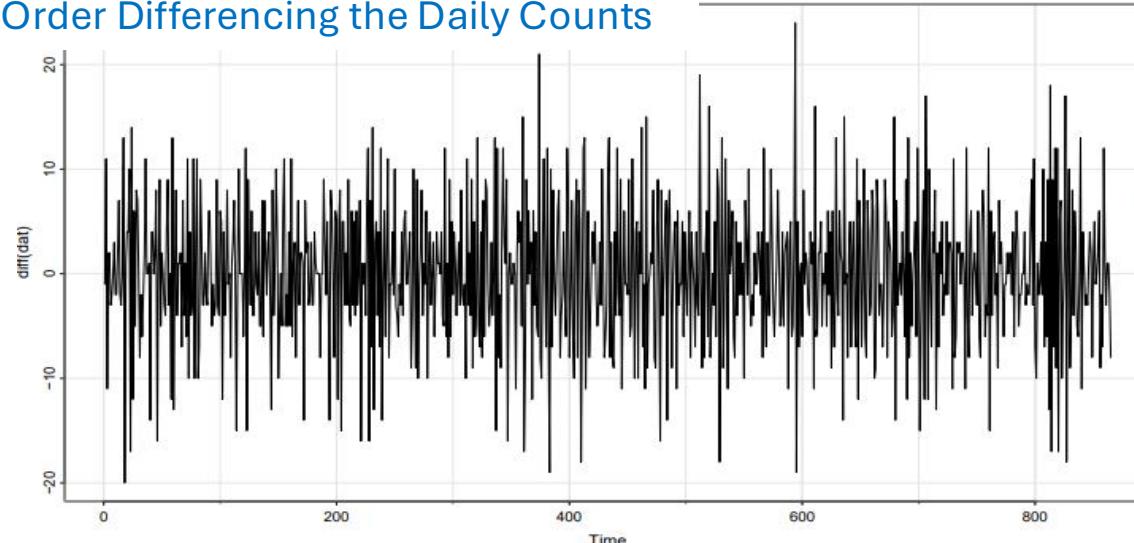
Community Policing Daily Counts Time Evolution



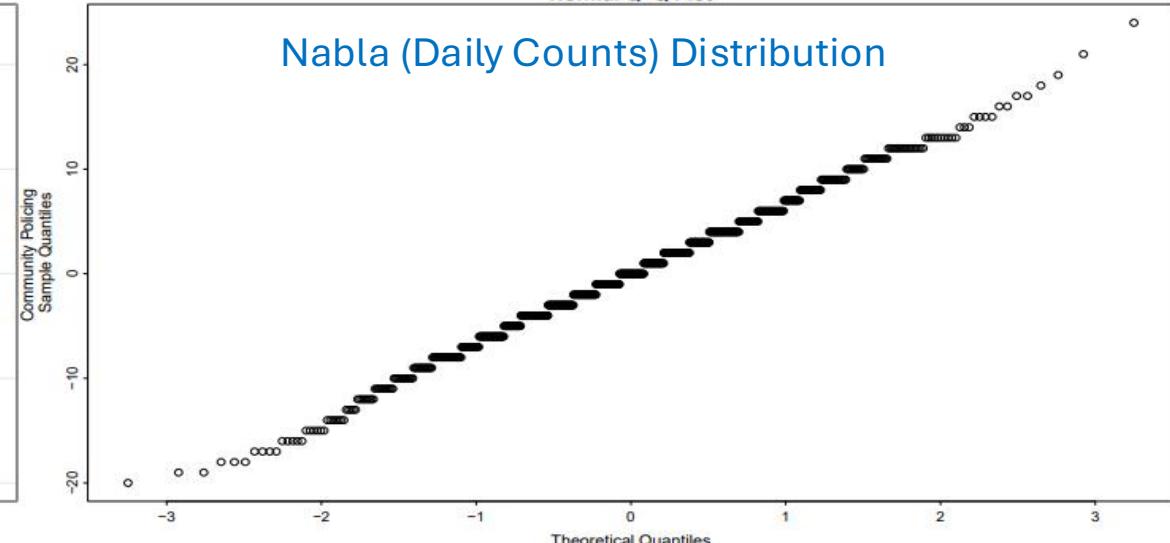
Community Policing Daily Counts Distribution



First Order Differencing the Daily Counts



Nabla (Daily Counts) Distribution



Daily Community Policing Diagnostics

Nabla (Counts) Model

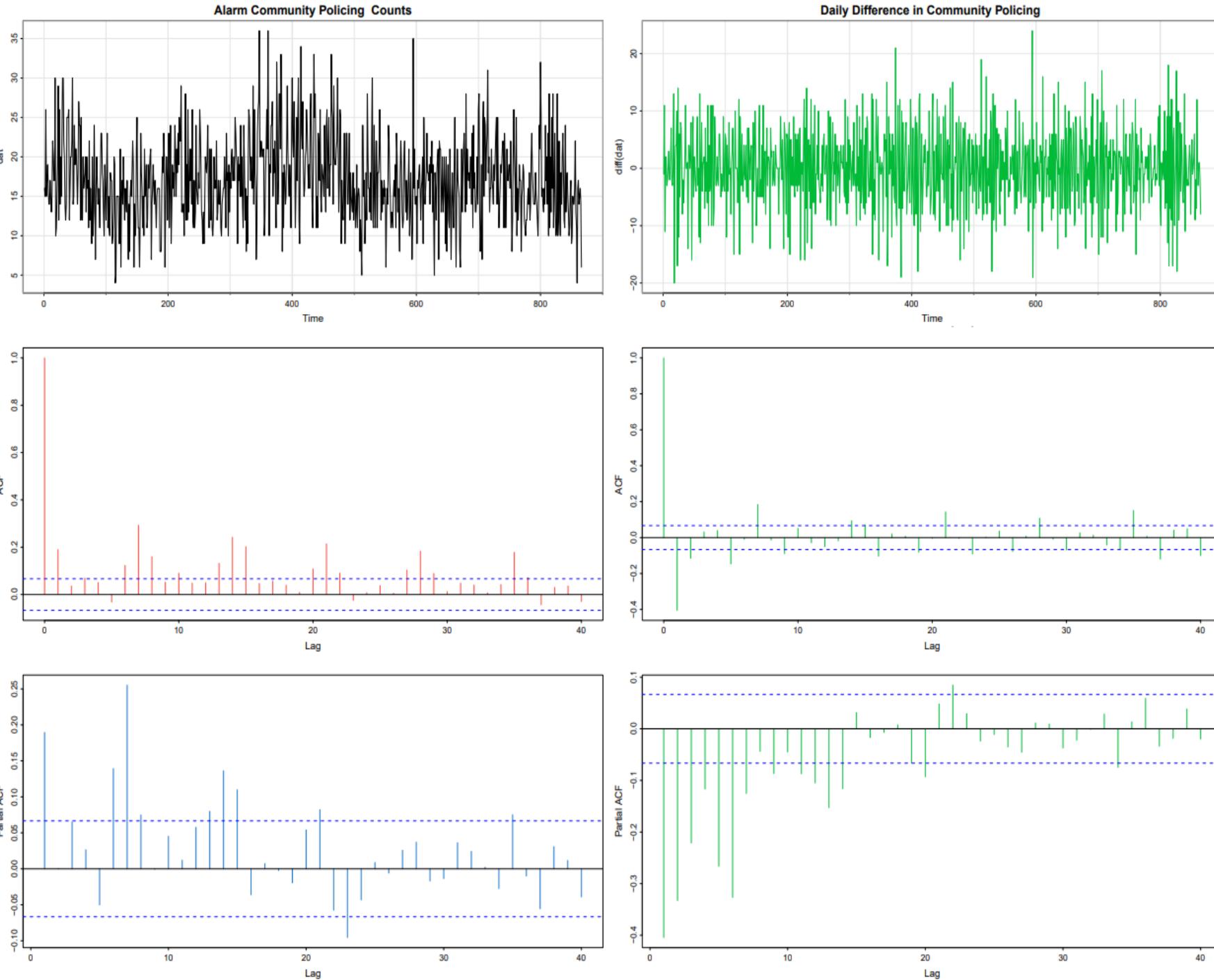
SARIMA (**diff(counts)**)

$p=0, d=0, q=2$

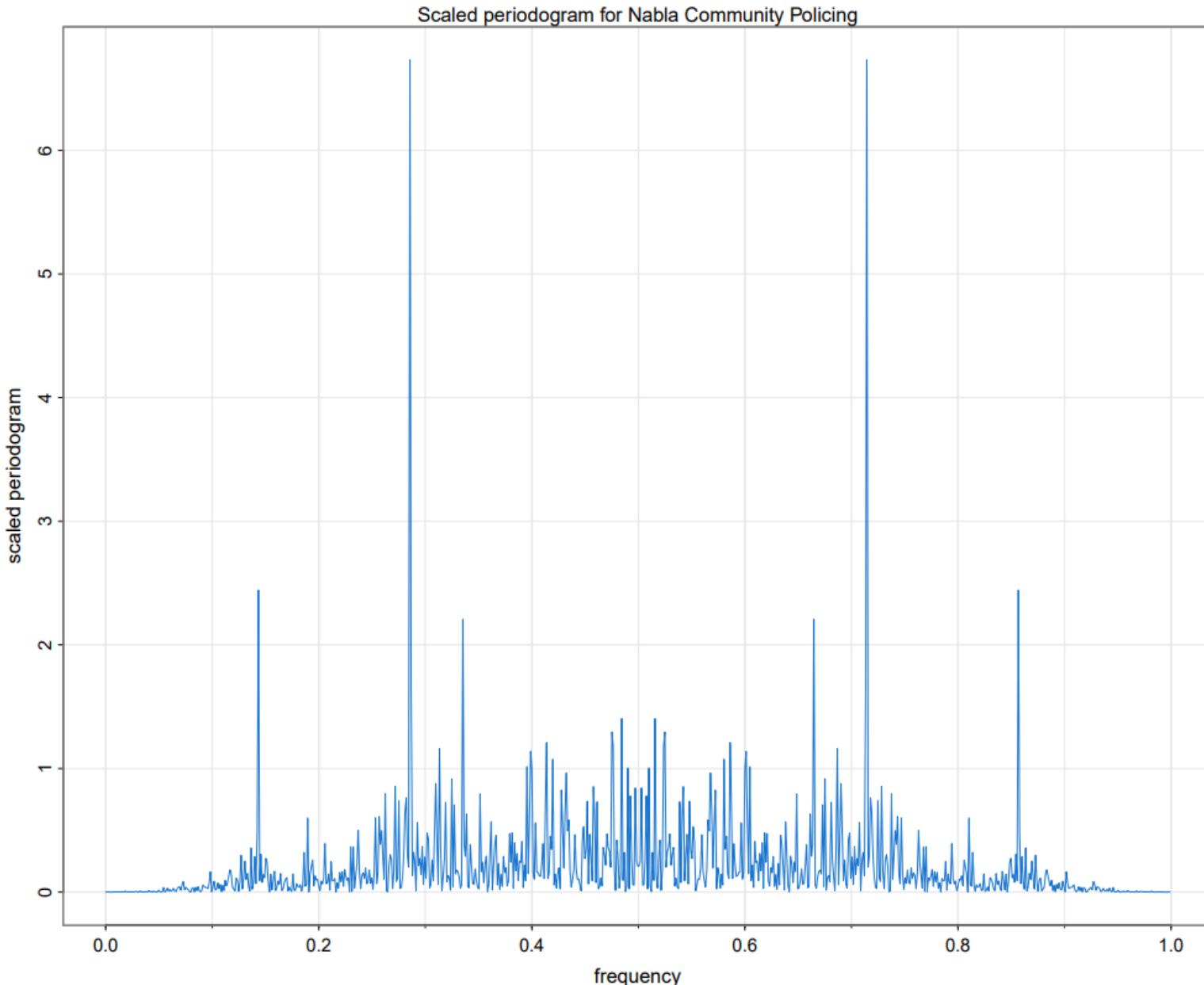
$P=3, D=0, Q=0$

7-day seasonality

$S=7$



Daily Growth Periodogram for Community Policing Incidents



Daily Community Policing Model Goodness of Fit

Community Policing
Day Diff(Count) Model

SARIMA (**diff(counts)**)

p=0, d=0, q=2

P=3, D=0, Q=0

7-day seasonality

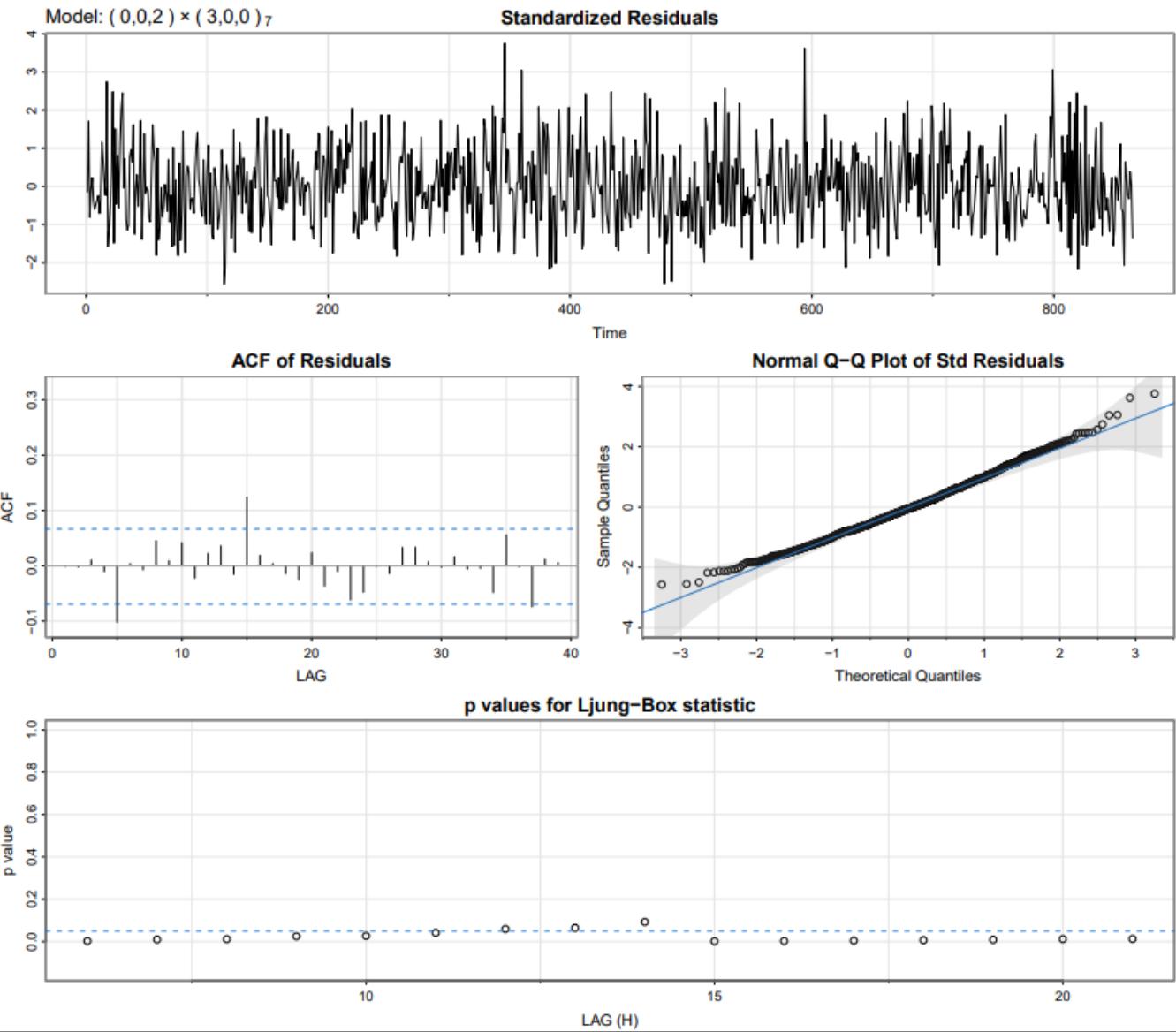
S=7

Coefficients:

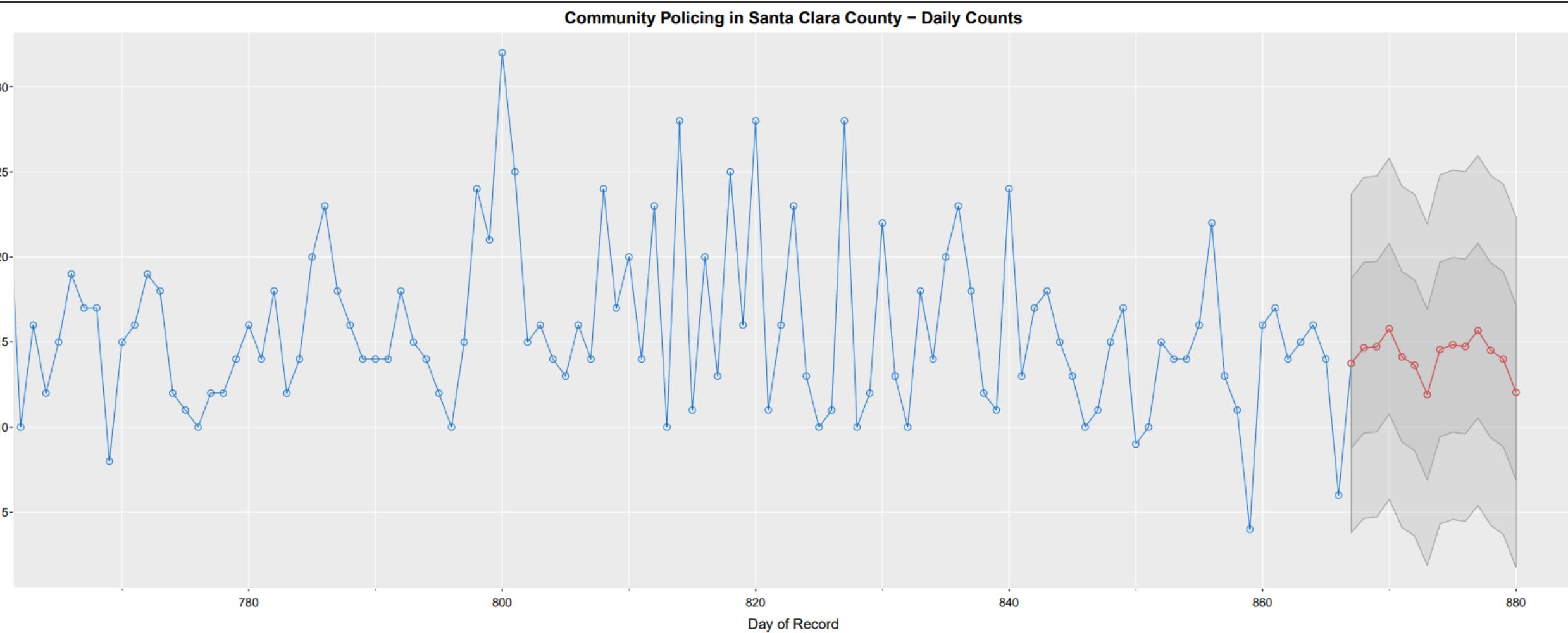
	Estimate	SE	t.value	p.value
ma1	-0.8845	0.0346	-25.5686	0.0000
ma2	-0.0894	0.0345	-2.5923	0.0097
sar1	0.1878	0.0347	5.4060	0.0000
sar2	0.1091	0.0349	3.1250	0.0018
sar3	0.1070	0.0343	3.1219	0.0019
xmean	-0.0035	0.0077	-0.4525	0.6510

σ^2 estimated as 24.75586 on 859 degrees of freedom

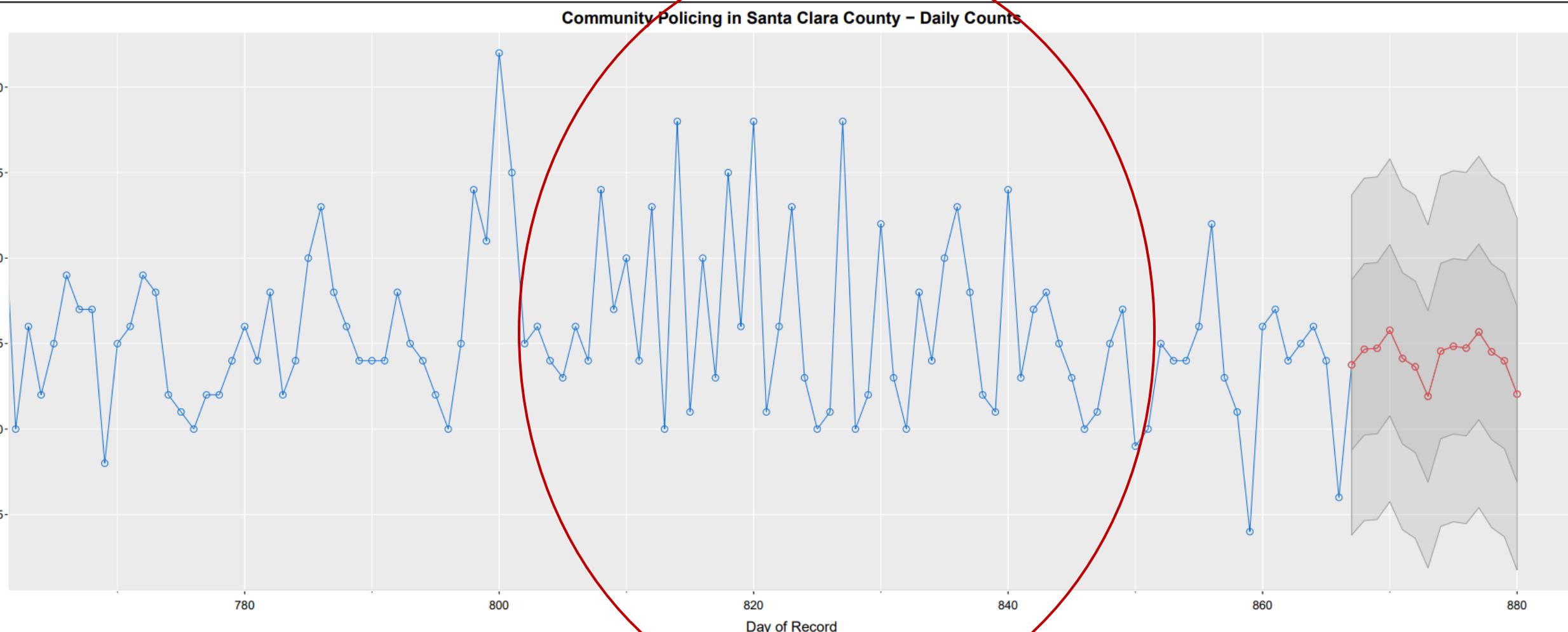
AIC = 6.066645 AICc = 6.066758 BIC = 6.105187



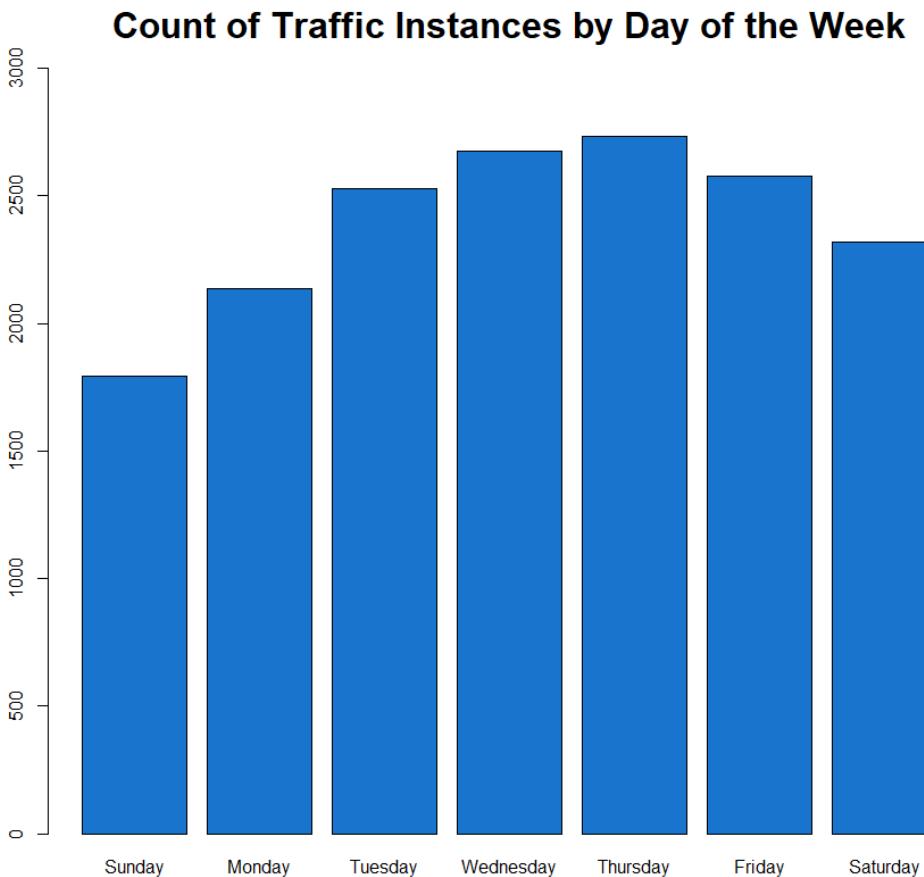
Community Policing 14-Day Prediction for Dec. 2019



Community Policing 14-Day Prediction for Dec. 2019



Daily Trends: Traffic



- Analysis Focus

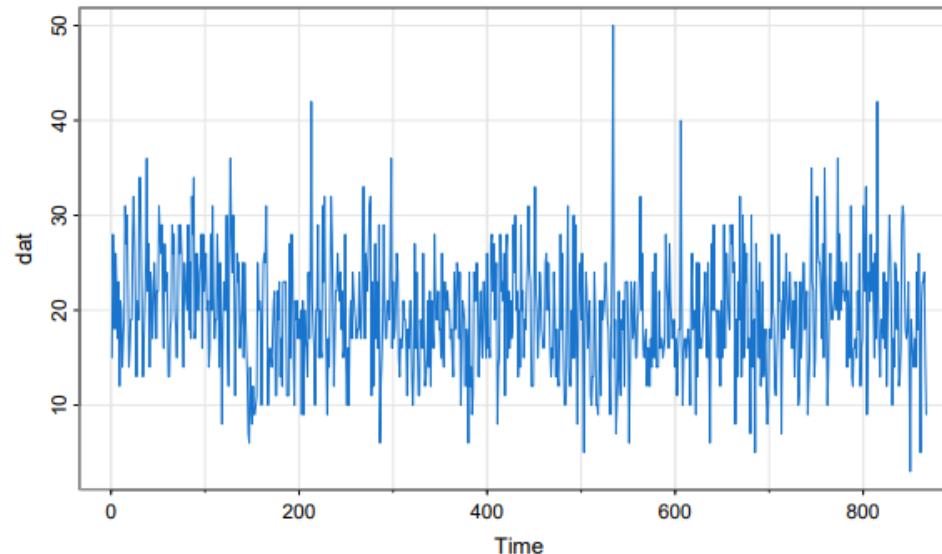
on preventive policing incidents, each with more than 10 counts daily

- Disorder - total recorded 16,768
~8.5% of the total data

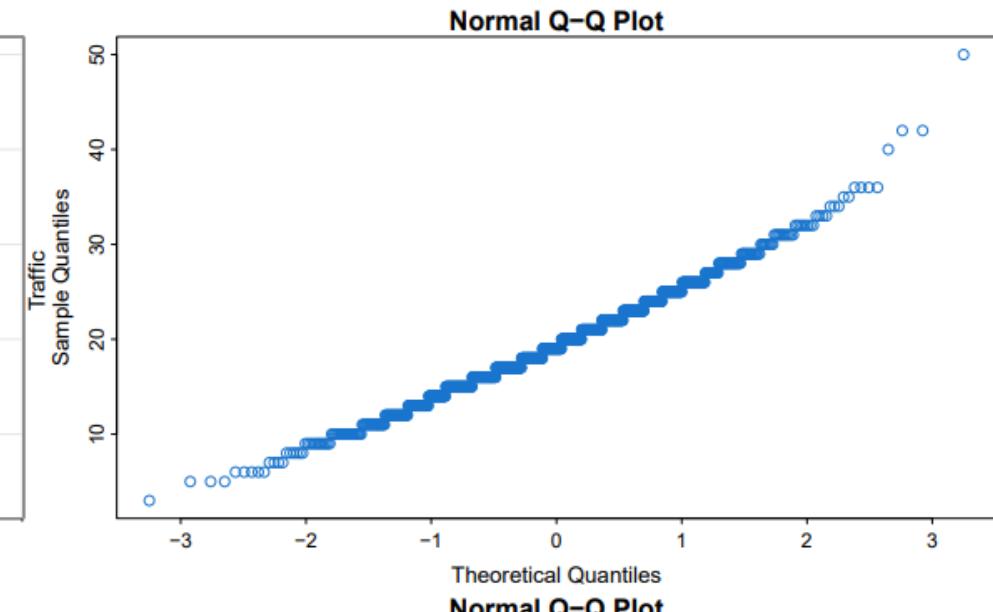
Daily Average: 20

Daily Traffic Incident Counts Diagnostics

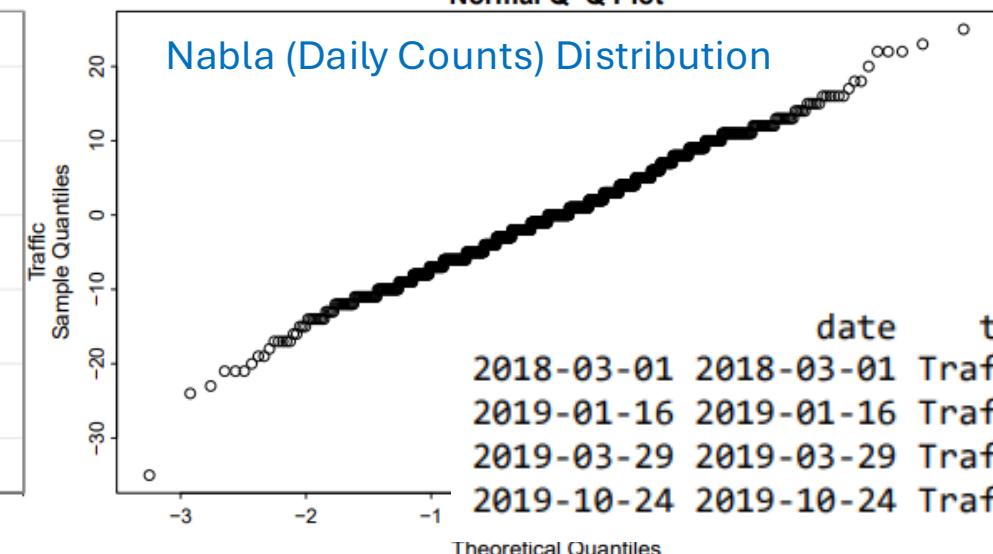
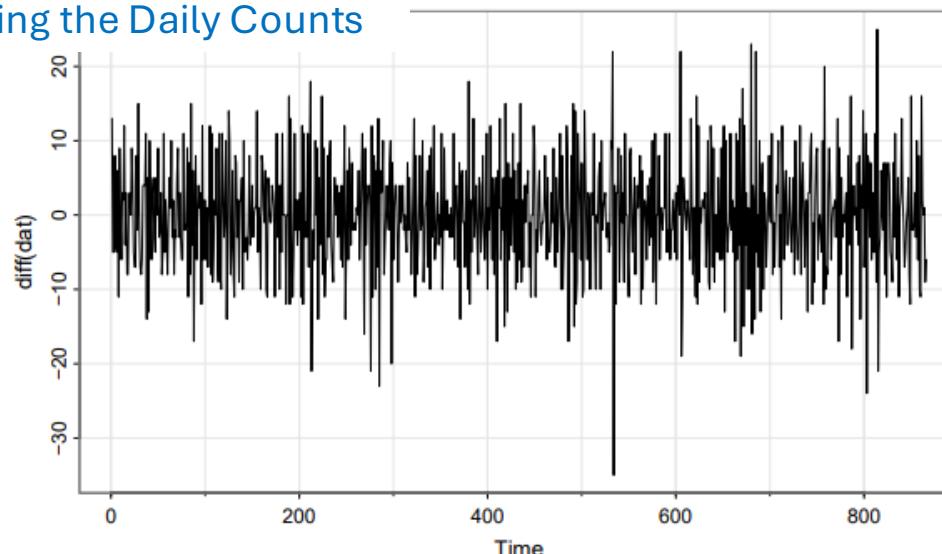
Traffic Daily Counts Time Evolution



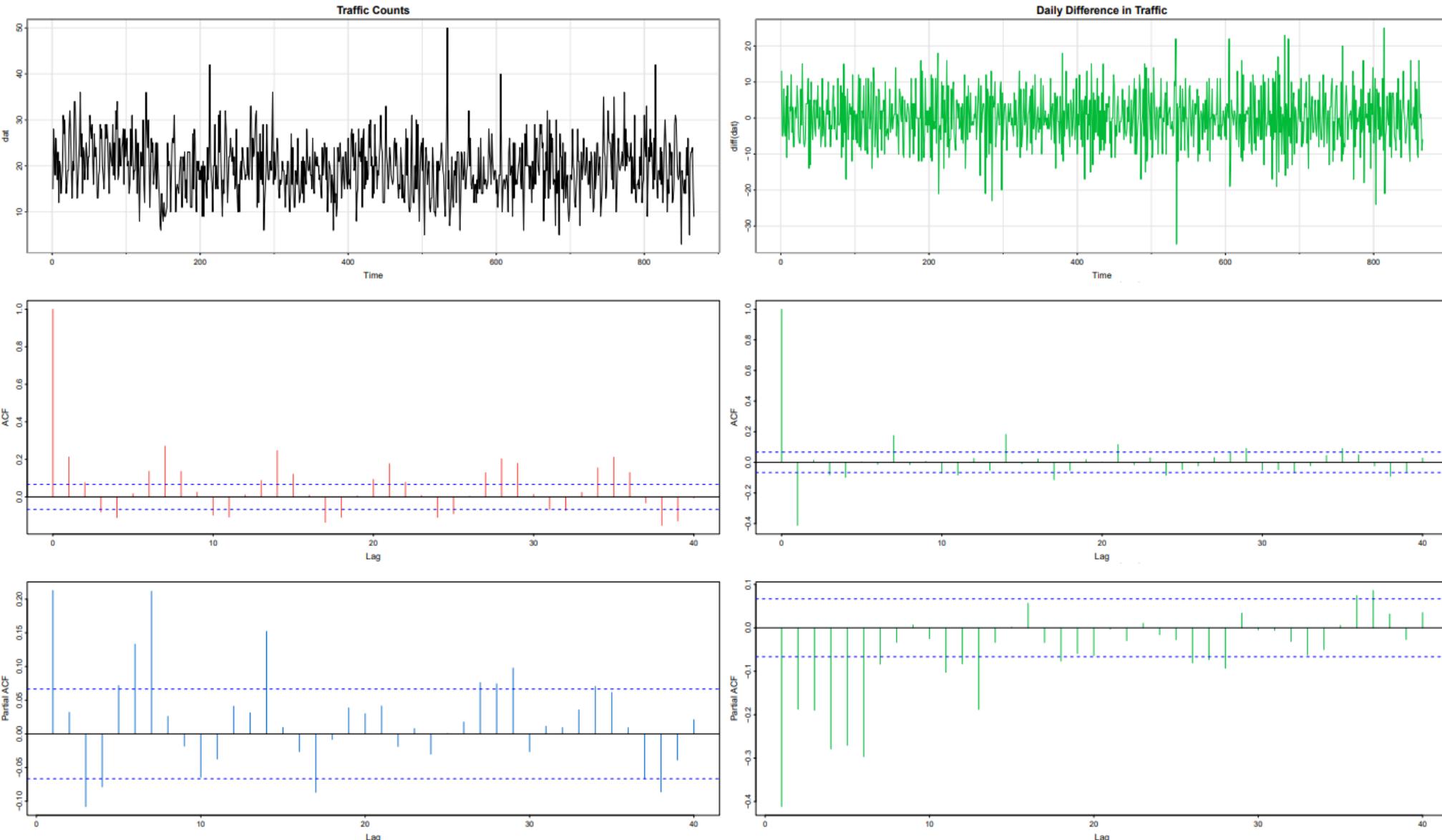
Traffic Daily Counts Distribution



First Order Differencing the Daily Counts



Daily Traffic Diagnostics



Counts Model

SARIMA (counts)

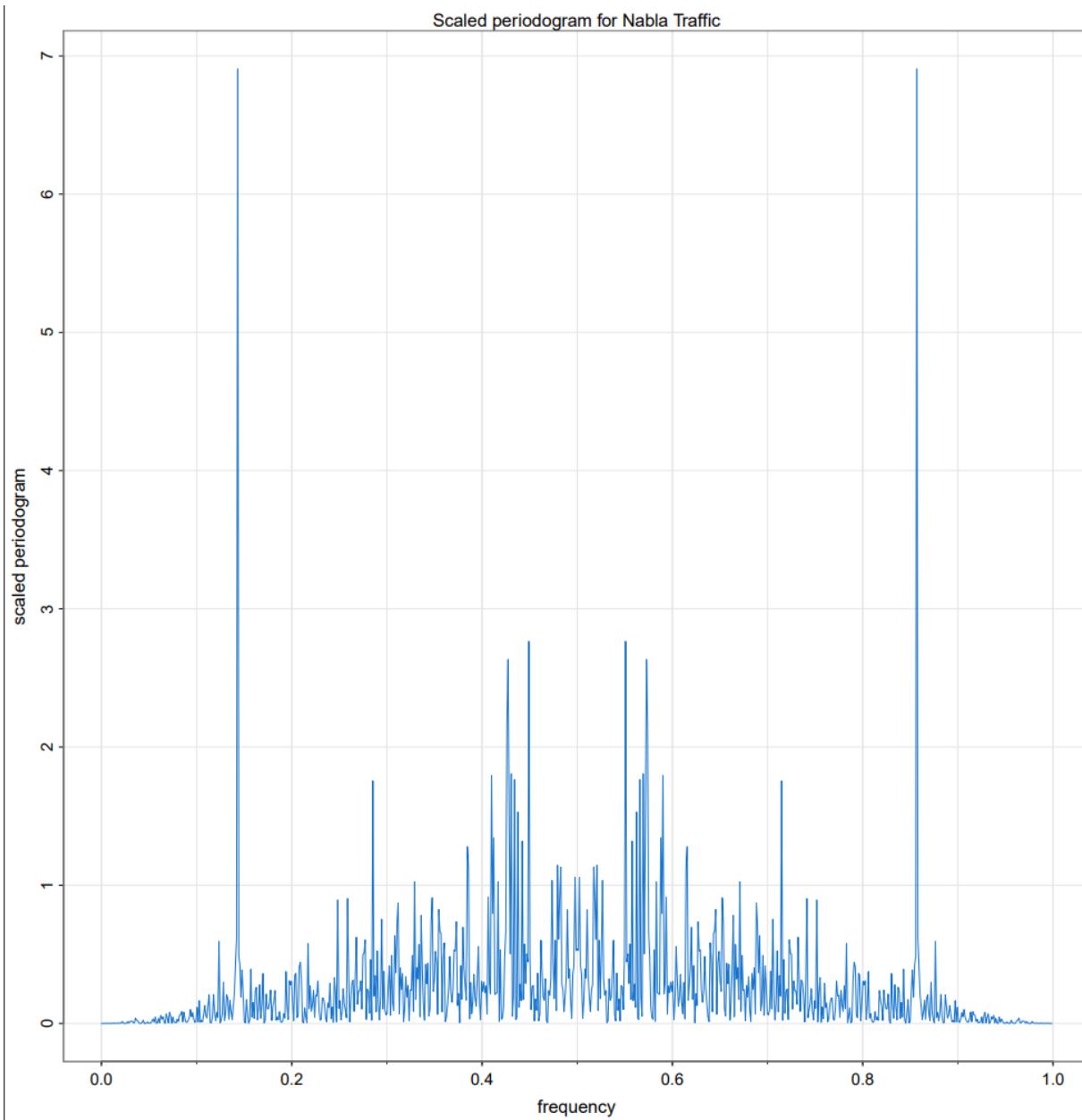
$p=2, d=0, q=1$

$P=1, D=0, Q=0$

7-day seasonality

$S=7$

Daily Growth Periodogram for Traffic Incidents



Daily Traffic Model Goodness of Fit

Traffic
Day Count Model

SARIMA (counts)

p=1, d=0, q=0

P=2, D=0, Q=0

7-day seasonality

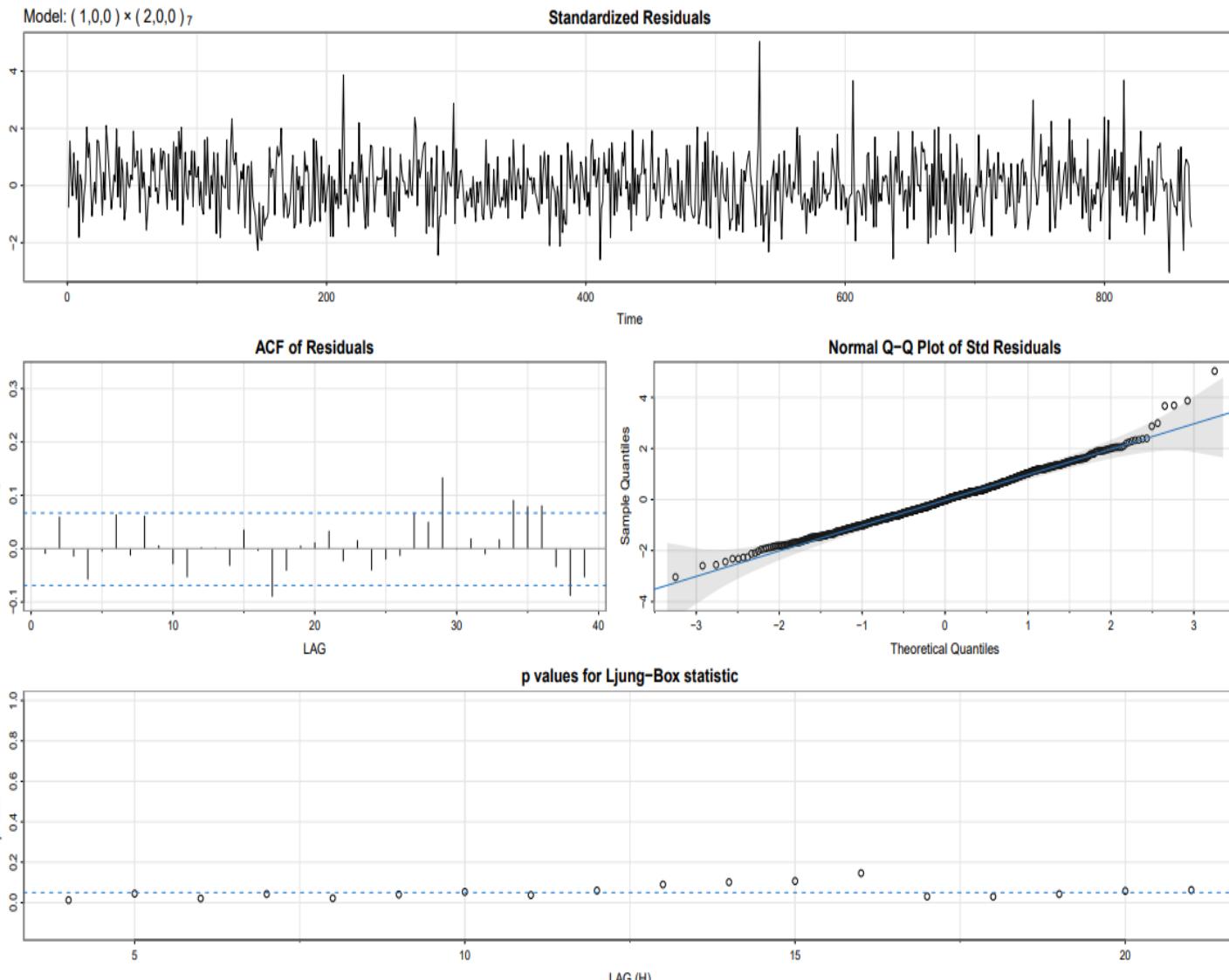
S=7

Coefficients:

	Estimate	SE	t.value	p.value
ar1	0.1650	0.0339	4.8656	0
sar1	0.2013	0.0338	5.9585	0
sar2	0.1811	0.0336	5.3824	0
xmean	19.5405	0.3681	53.0831	0

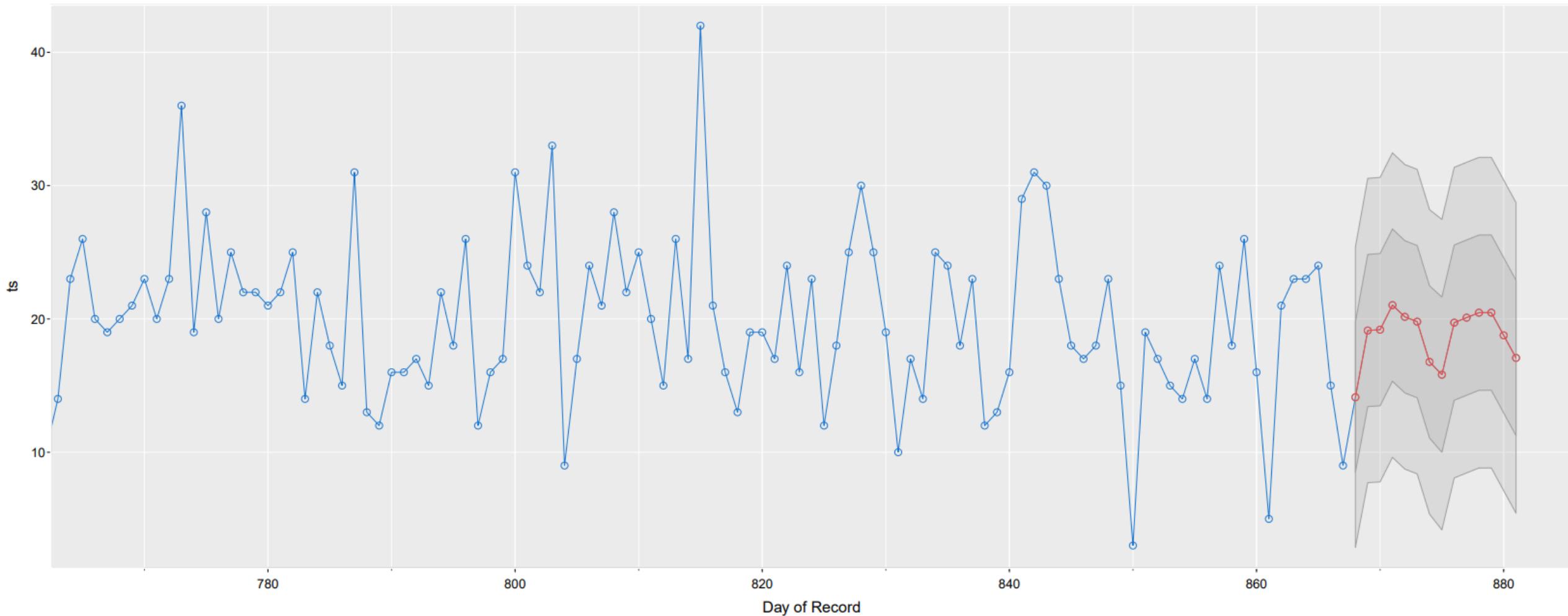
σ^2 estimated as 31.72 on 863 degrees of freedom

AIC = 6.307432 AICc = 6.307485 BIC = 6.334912



Traffic 14-Day Prediction for Dec. 2019

Traffic Incidents in Santa Clara County – Daily Counts



ARMA Model Formulation for Weekly Counts

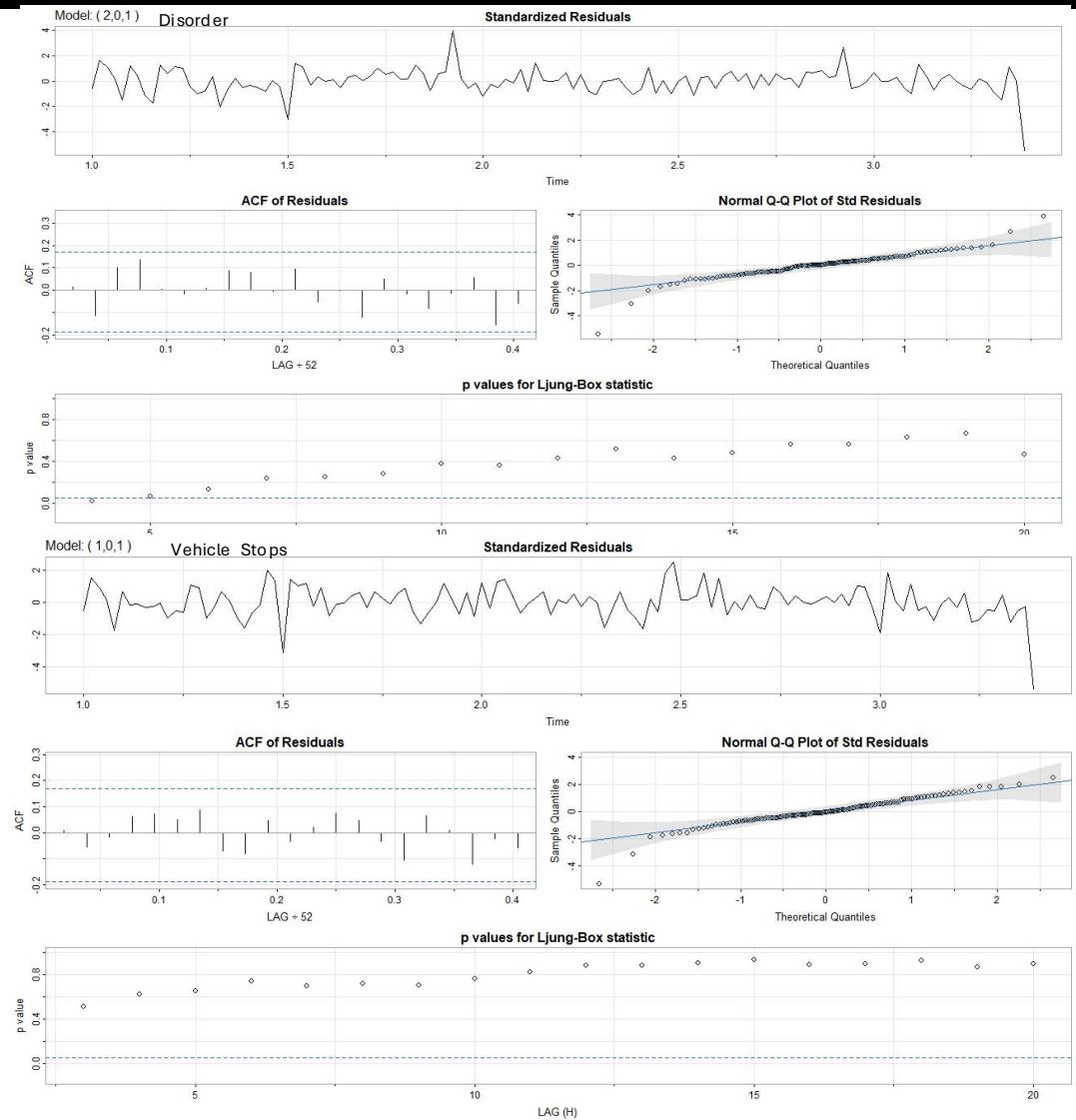
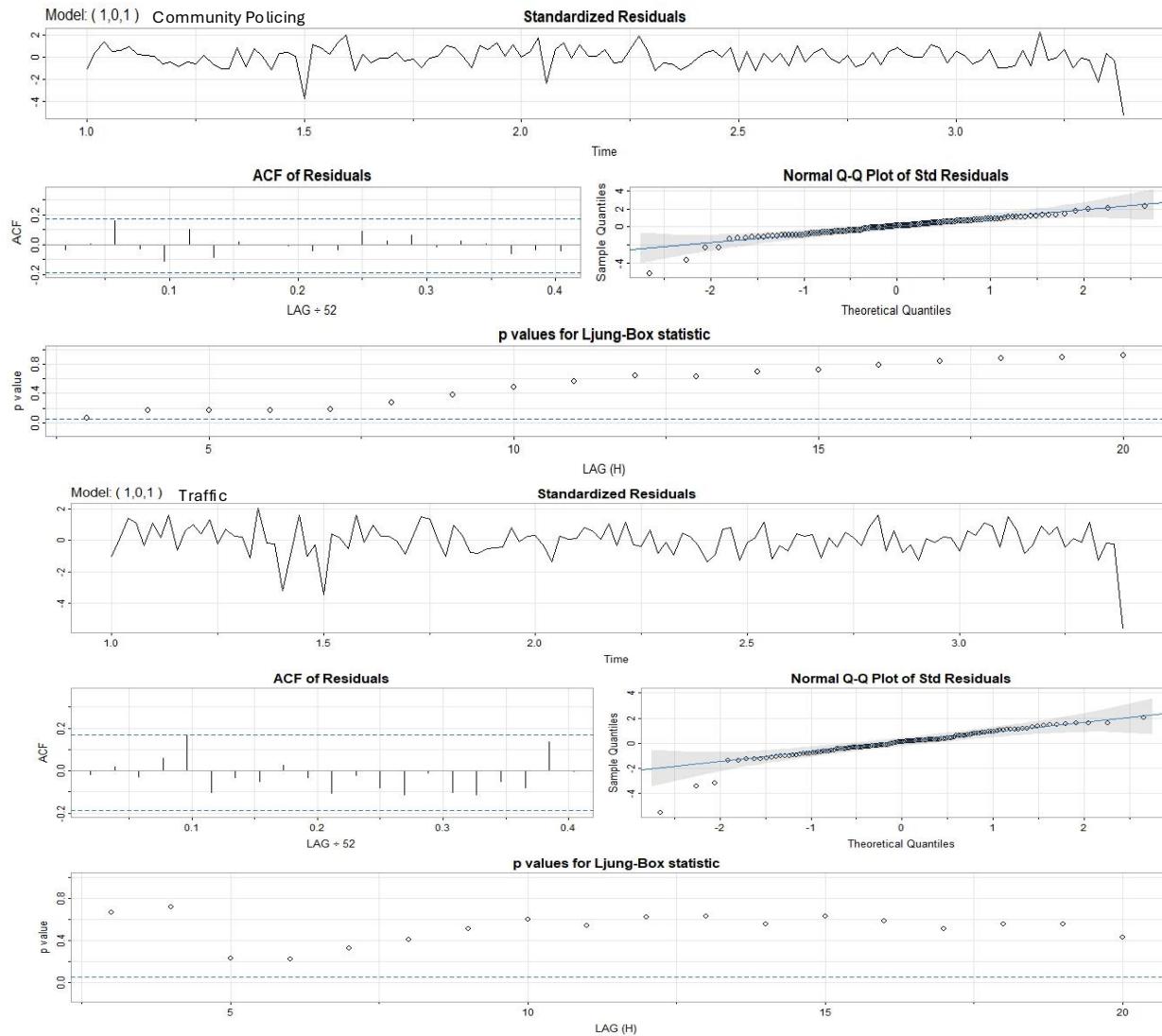
- **Community Policing**
 - ARMA(1, 1)
- **Disorder**
 - ARMA (2,1)
- **Traffic**
 - ARMA(1,1)
- **Vehicle Stop**
 - ARMA(1,1)

4. ARMA MODELS

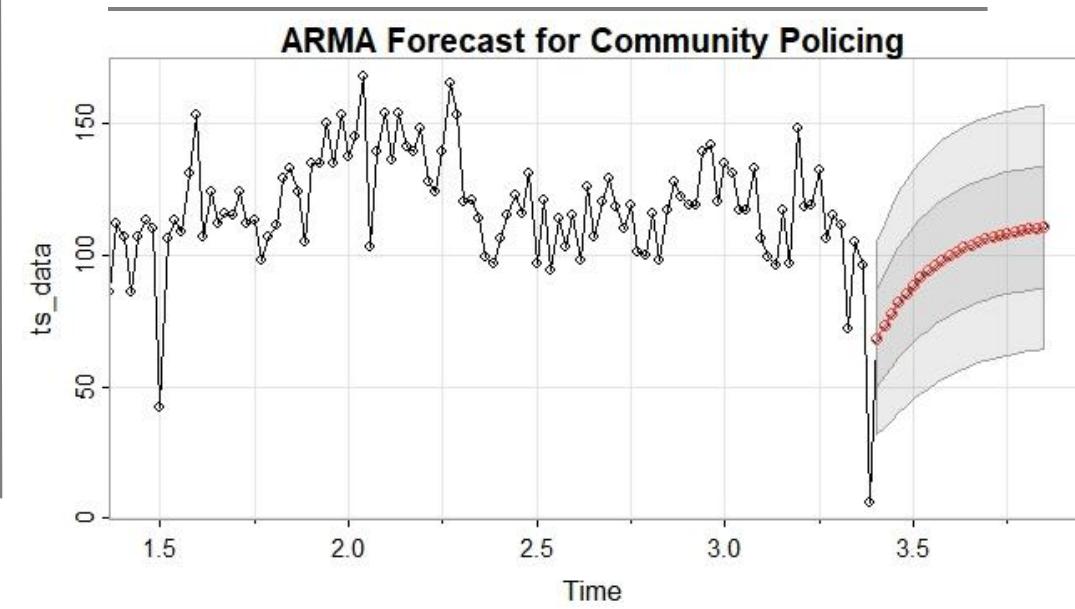
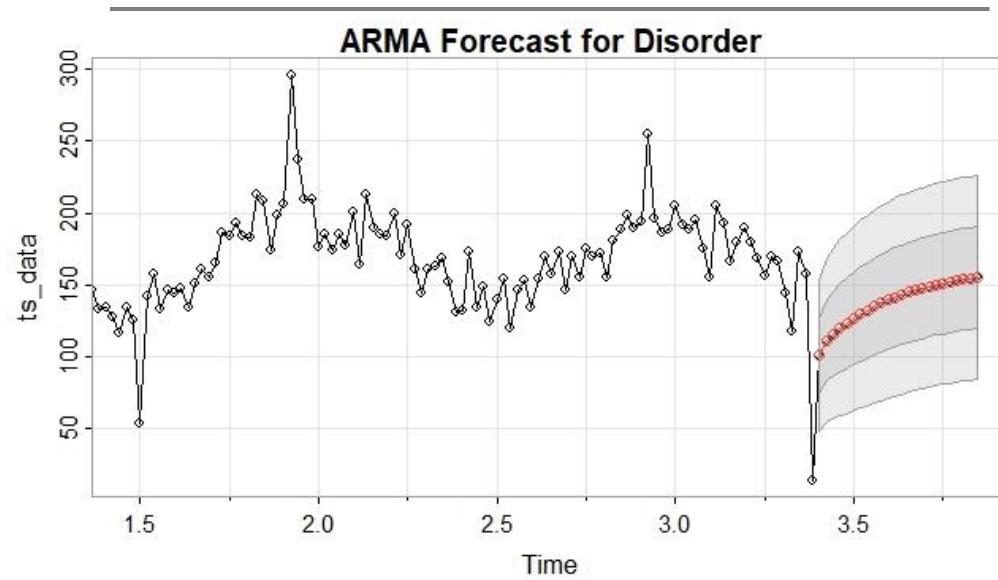
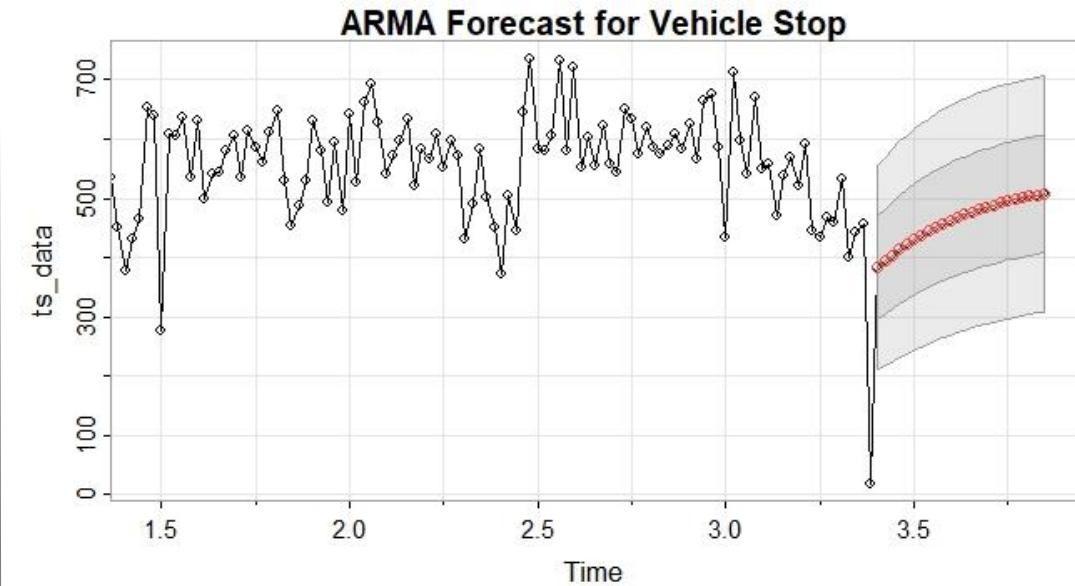
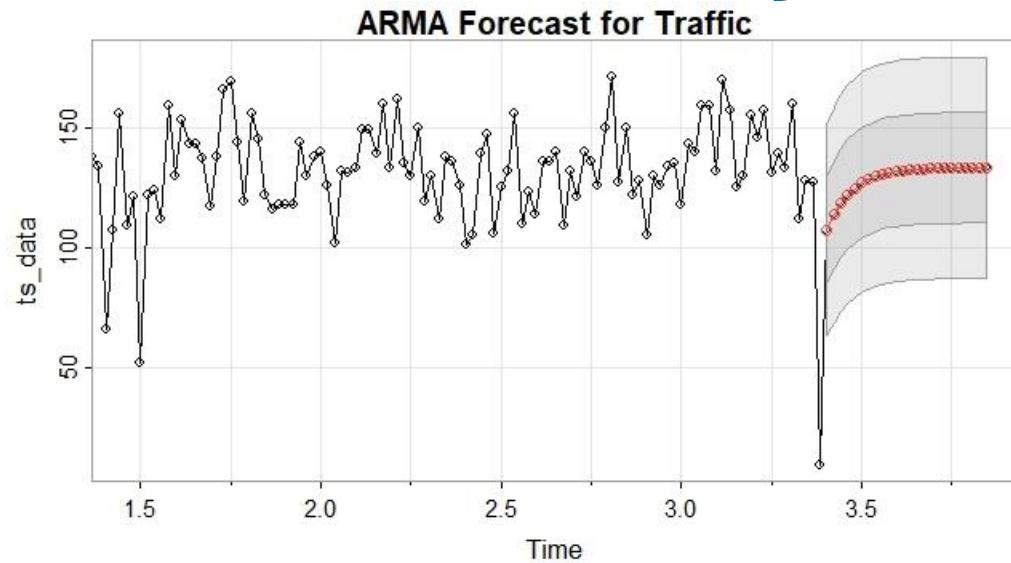
Table 4.1 *Behavior of the ACF and PACF for ARMA Models*

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Weekly Counts Model Diagnostics

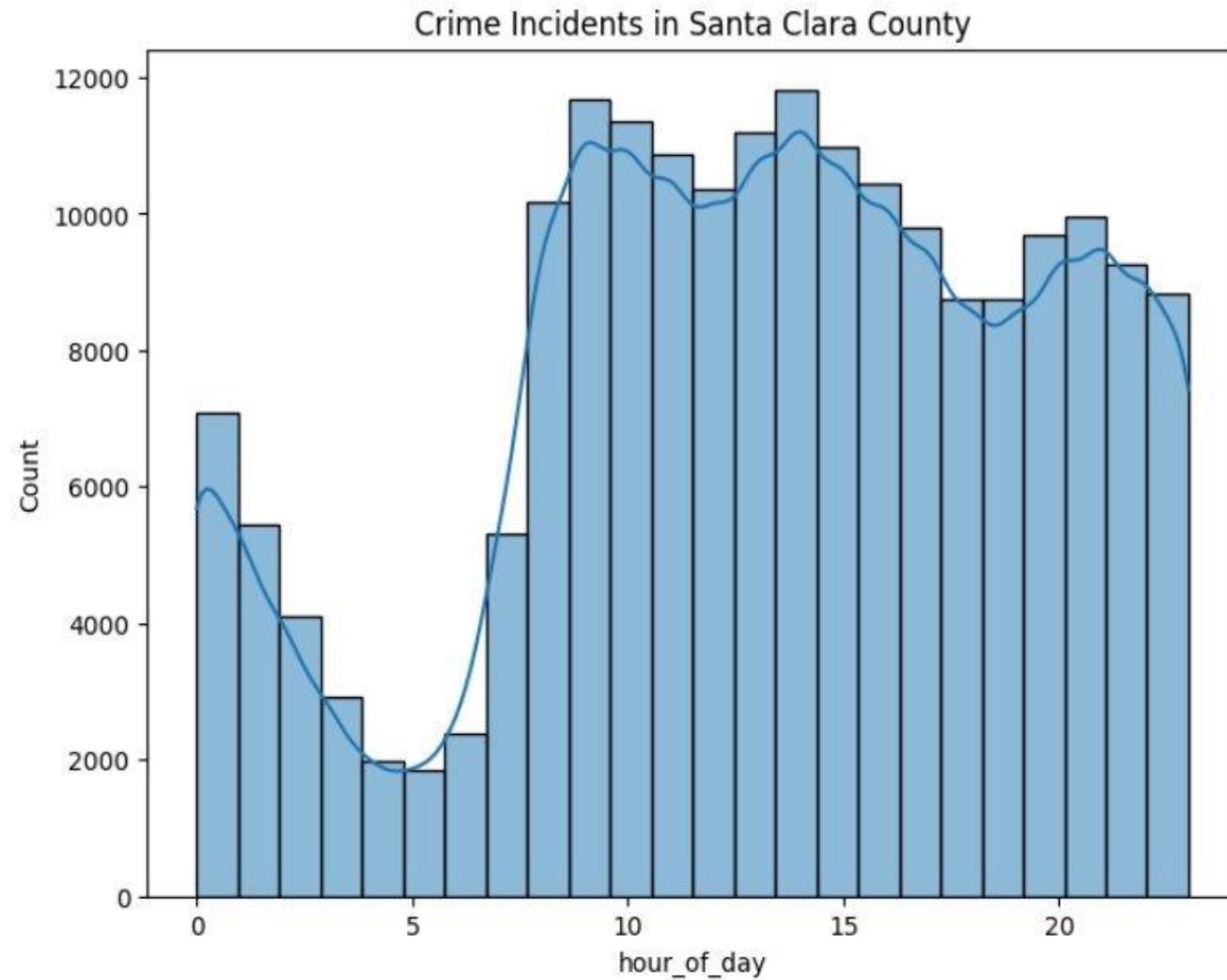


Weekly Counts Forecast Plots

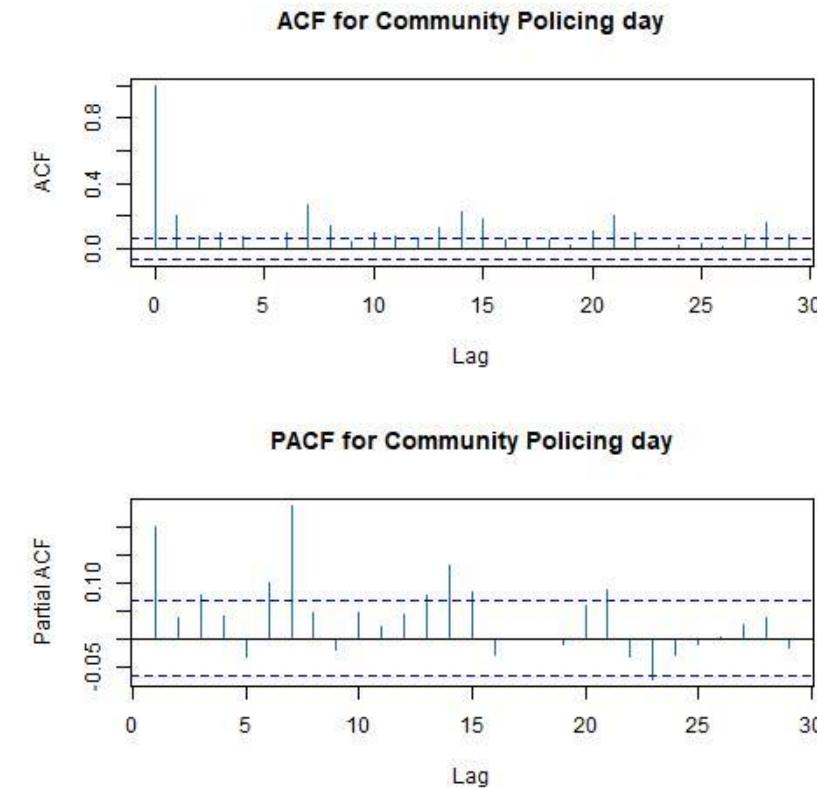
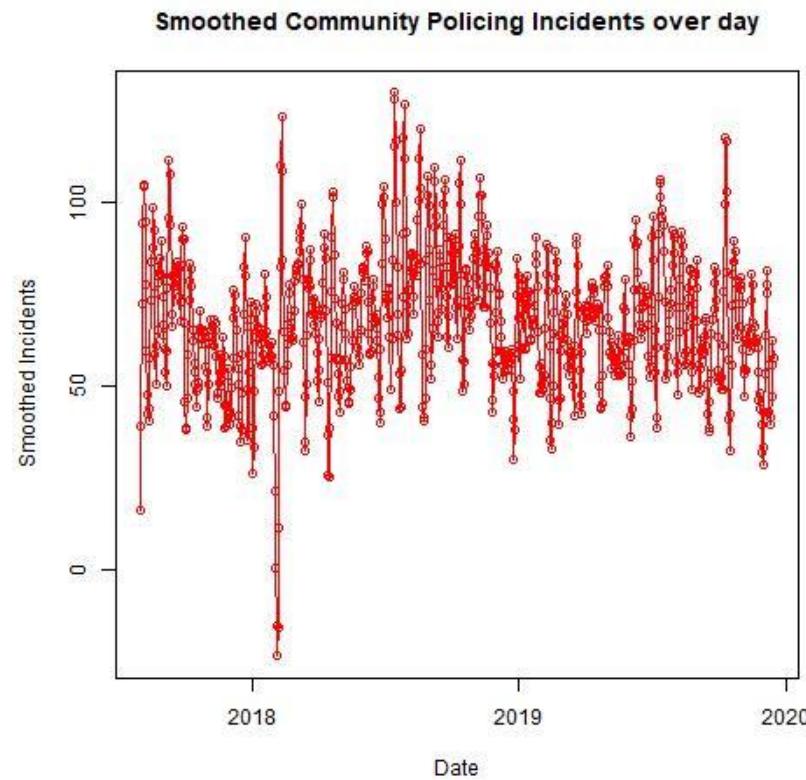


Santa Clara County Sheriff's Office

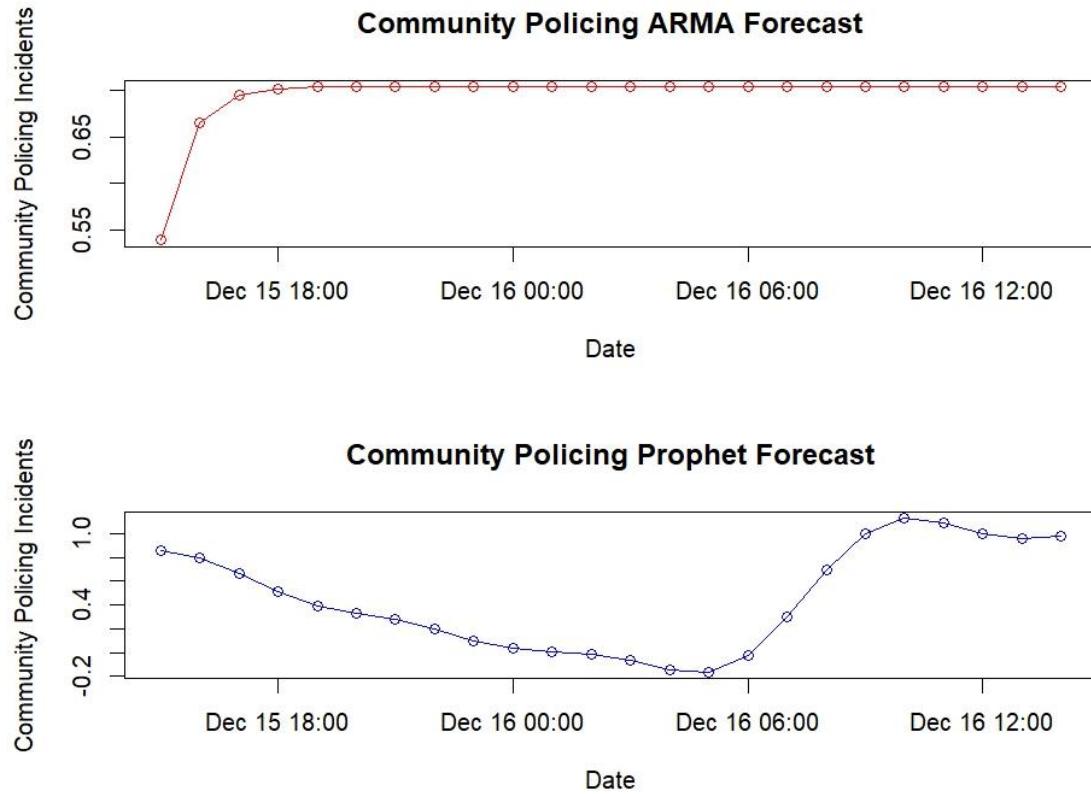
Hourly Incidents



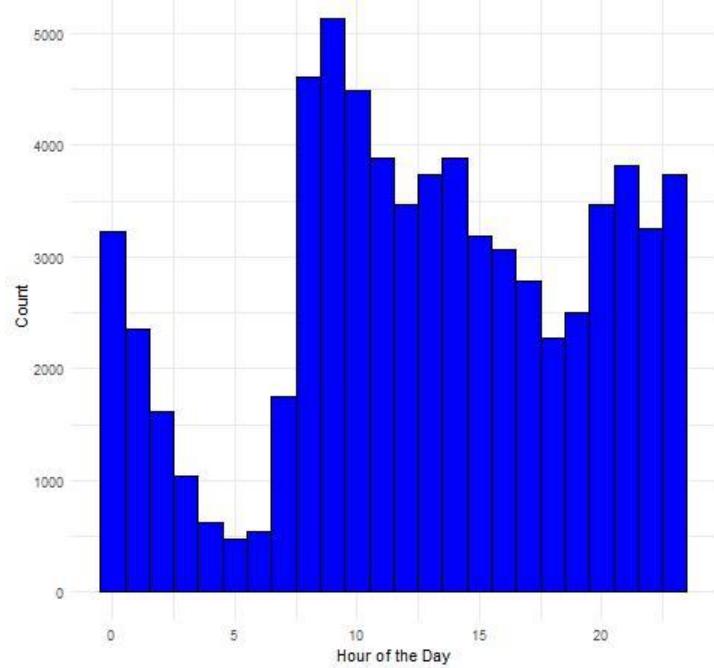
Hourly Analysis Community Policing



Community Policing Forecast Hourly

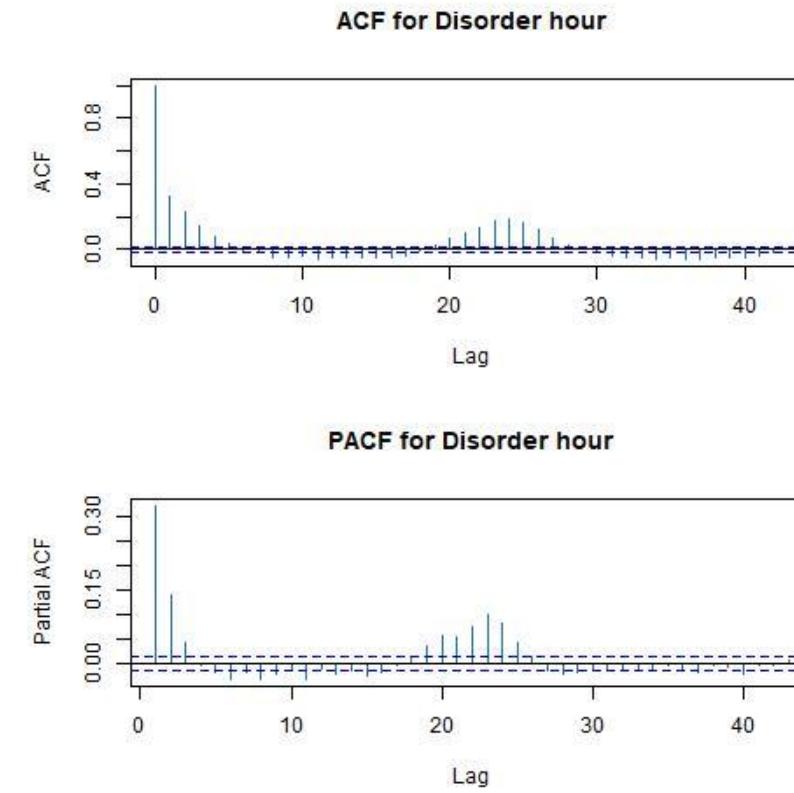
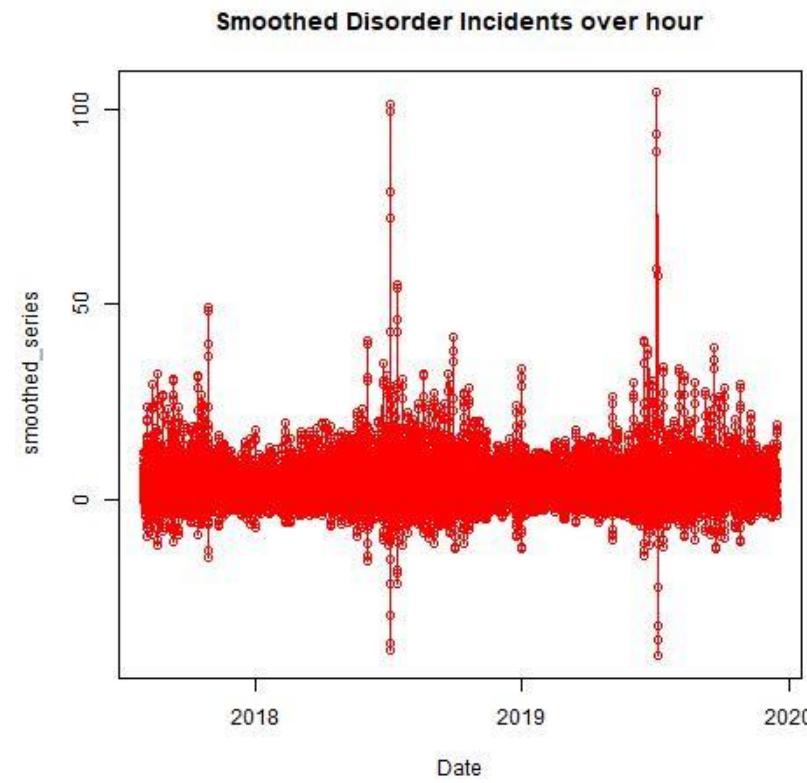


Histogram by Hour of the Day for Vehicle Stop

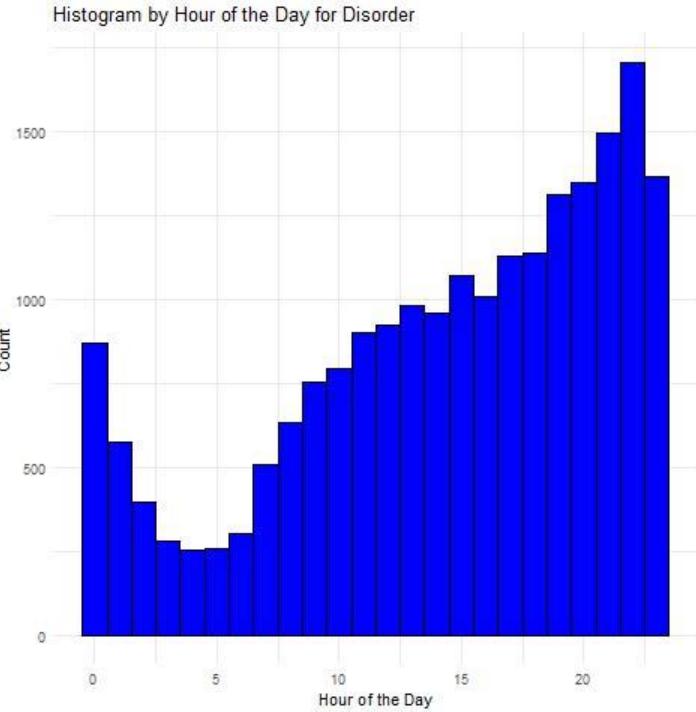
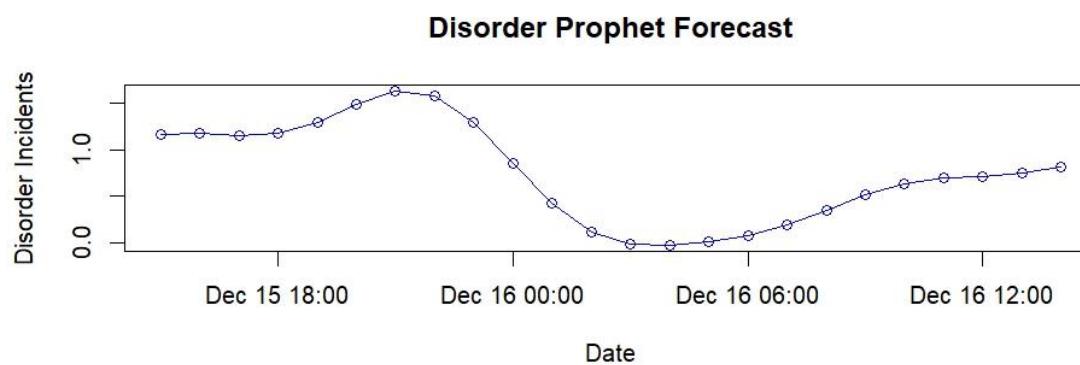
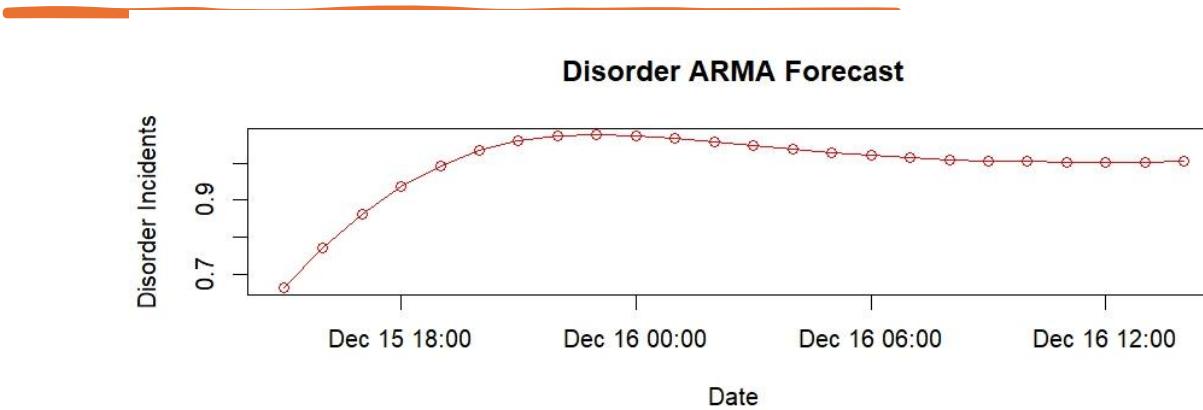


- ACF: The plot shows a significant spike at lag 1 and then rapidly decreases.
- PACF: The plot shows a significant spike at lag 1 and minor spikes thereafter.
- Model: AR(1)
 - A significant spike at lag 1 in both ACF and PACF suggests an AR(1) process.

Hourly Analysis Disorder

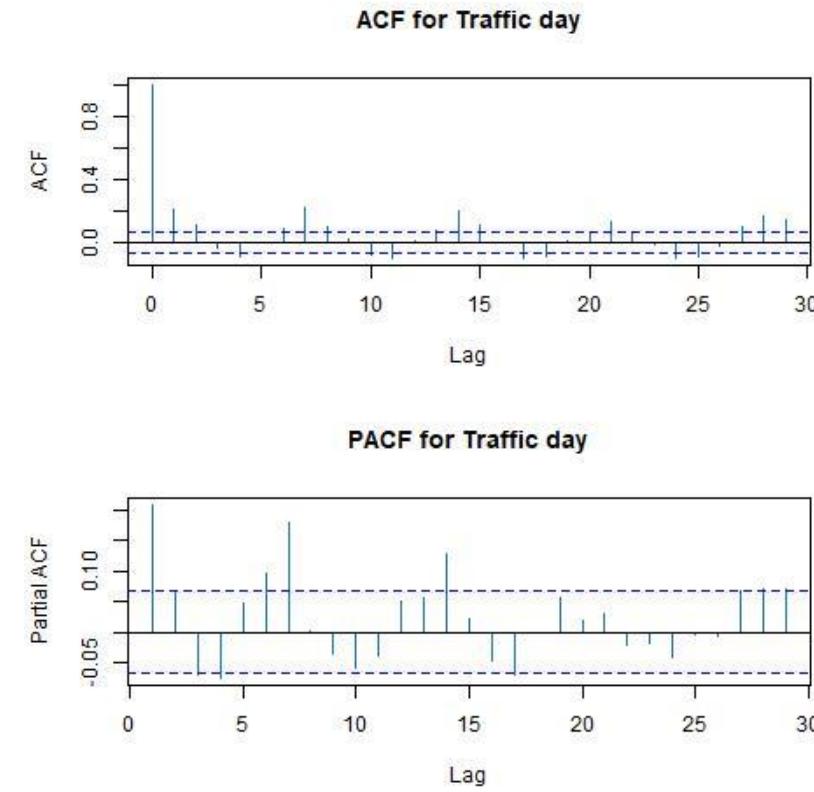
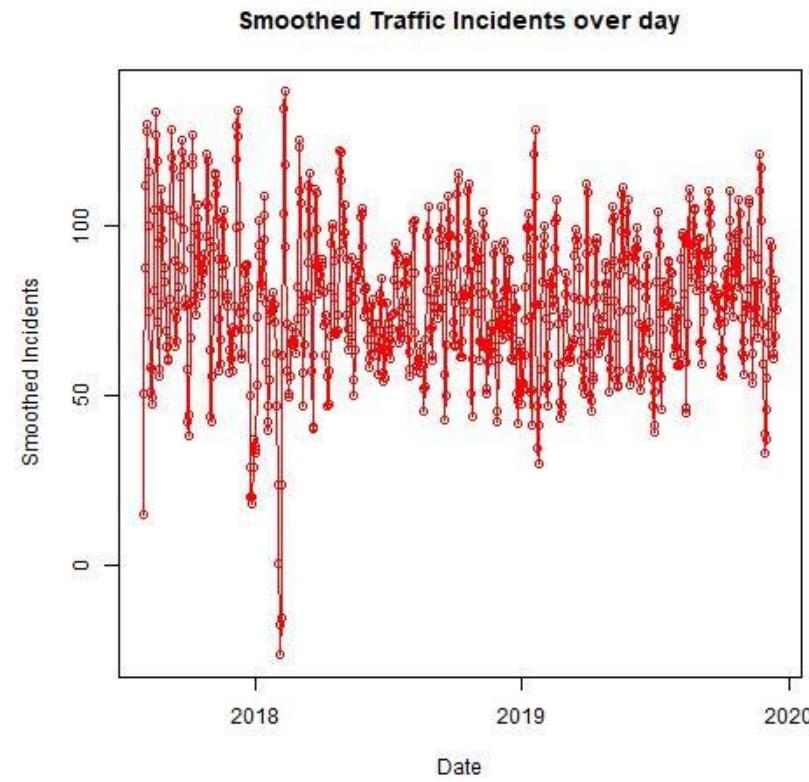


Disorder Forecast Hourly

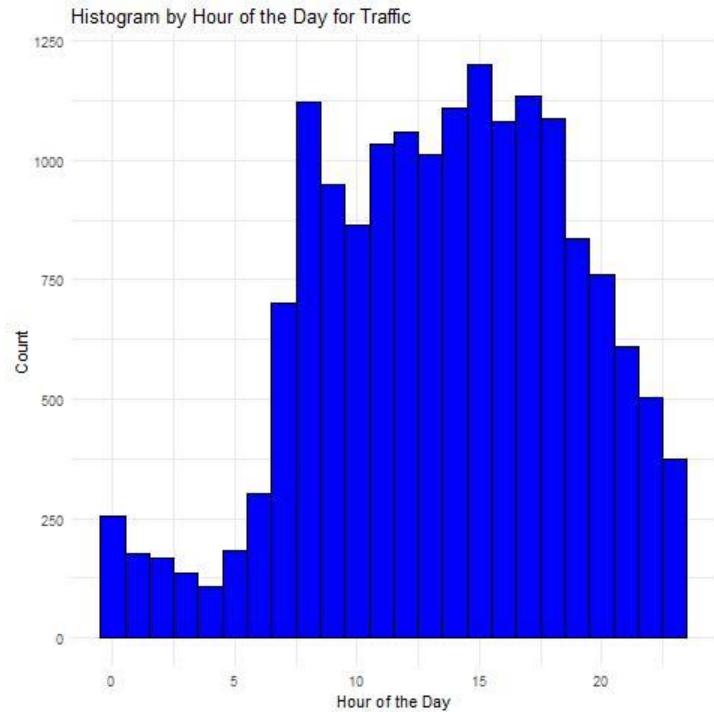
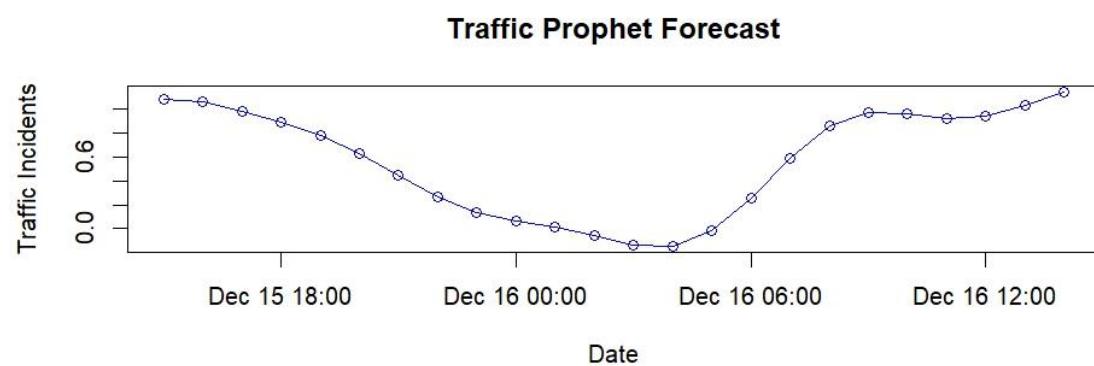
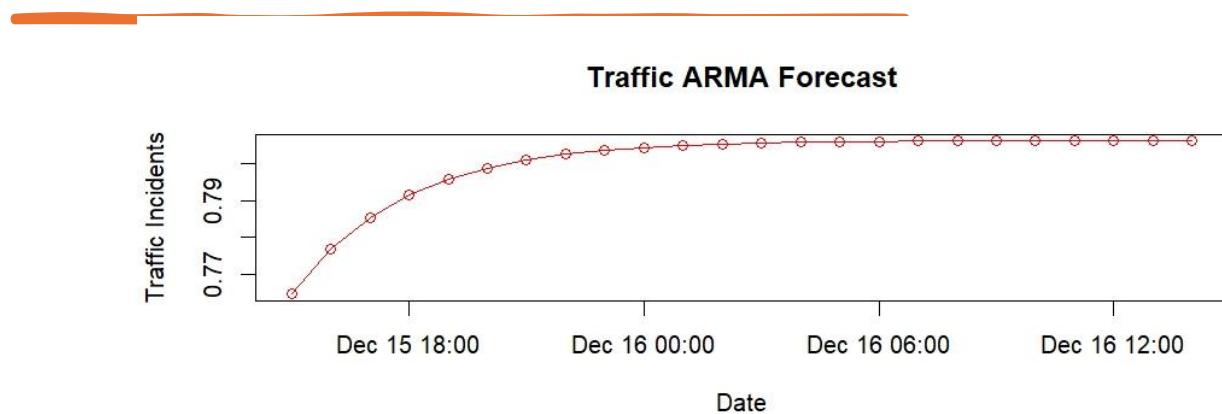


- ACF: Significant spikes at lags 1 and 24, with a gradual decay, indicating possible seasonality.
- PACF: Significant spikes at lags 1 and 24, suggesting a seasonal AR component.
- Model: ARMA 2,2
 - I was looking to keep it simple .

Hourly Analysis Traffic

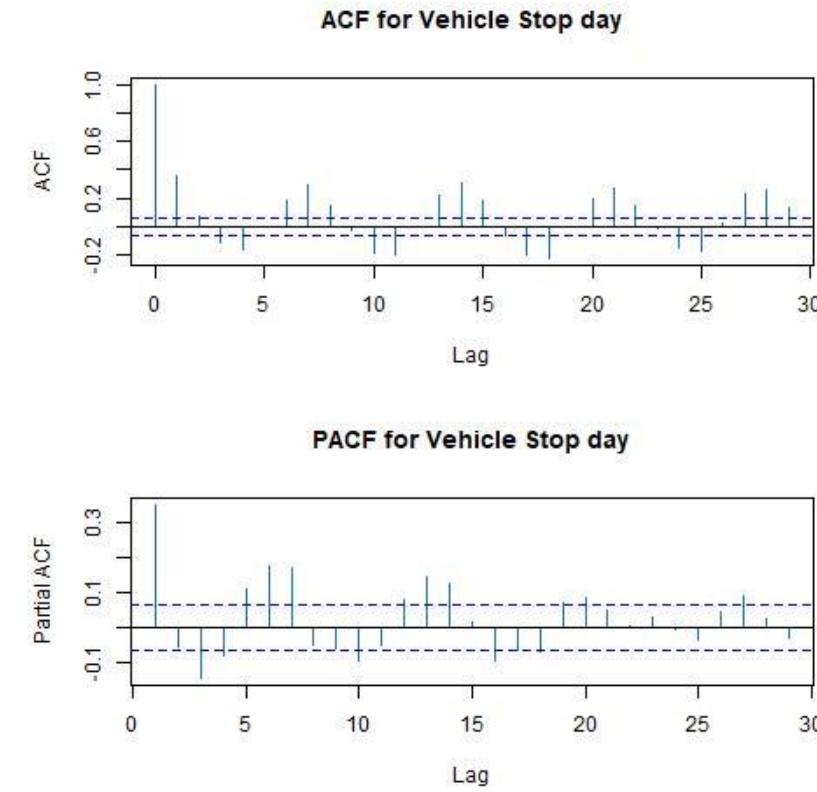
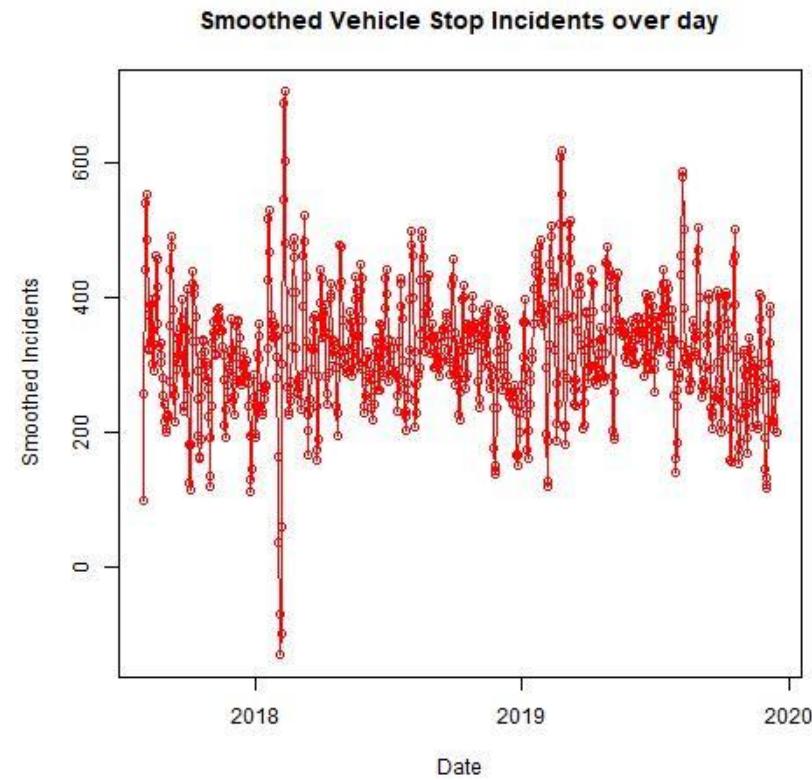


Traffic Forecast Hourly

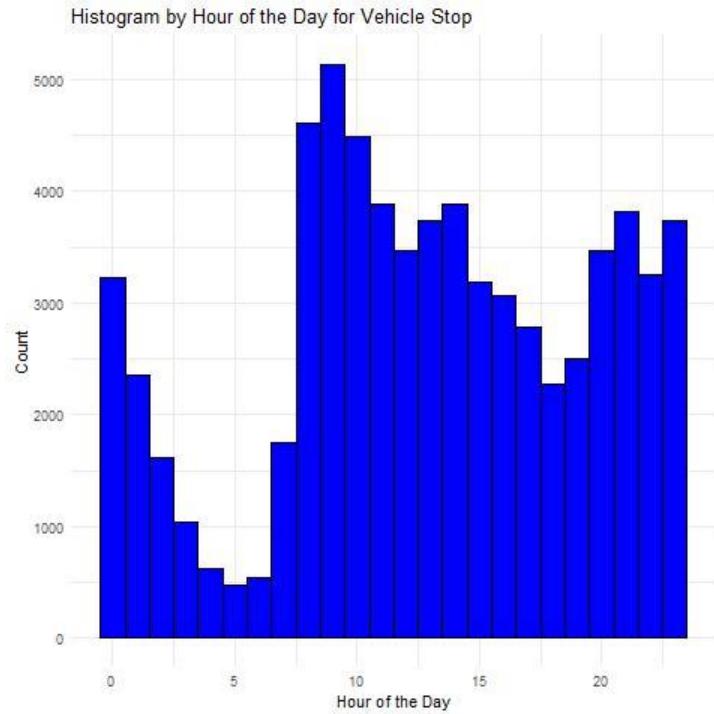
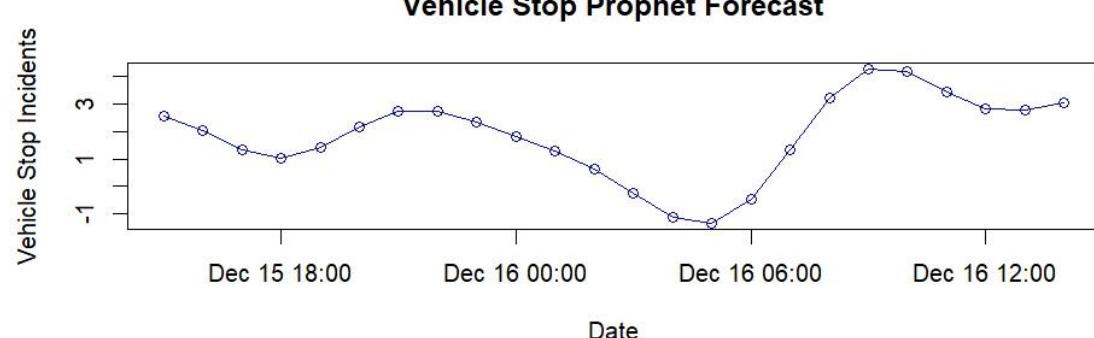
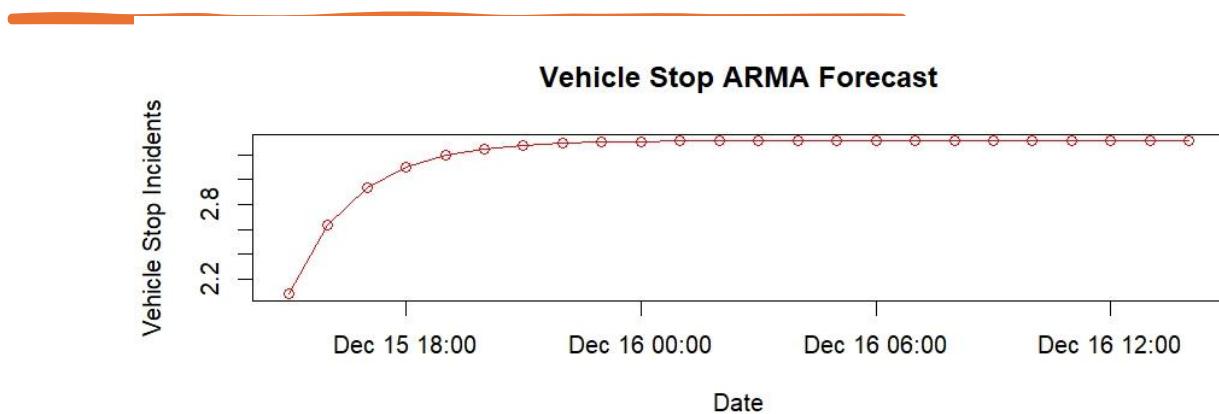


- ACF: Significant spike at lag 1, then minor spikes.
- PAC: Significant spike at lag 1 and then minor spikes.
- Model: AR(1)
 - The strong autocorrelation at lag 1 in both ACF and PACF indicates an AR(1) process.

Hourly Analysis Vehicle Stop



Vechicle Stop Forecast Hourly



- ACF: Shows a spike at lag 1 and decays slowly.
- PACF Significant spike at lag 1 and minor spikes thereafter.
- Model: AR(1)
 - The ACF plot's slow decay and PACF's significant spike at lag 1 indicate a potential AR(1) or ARMA(1,1) process.

Summary

- Data stationarity
- There is no increasing trend of the crime incident counts
- The daily incident counts have a seven-day cyclicity because of the city life workday / weekend pattern
- More than 90% of the sheriff's incident records are preventive routine
- Breaking the data into type of incidents allows meaningful modeling and forecasting



Questions?

- eene@tamu.edu
-