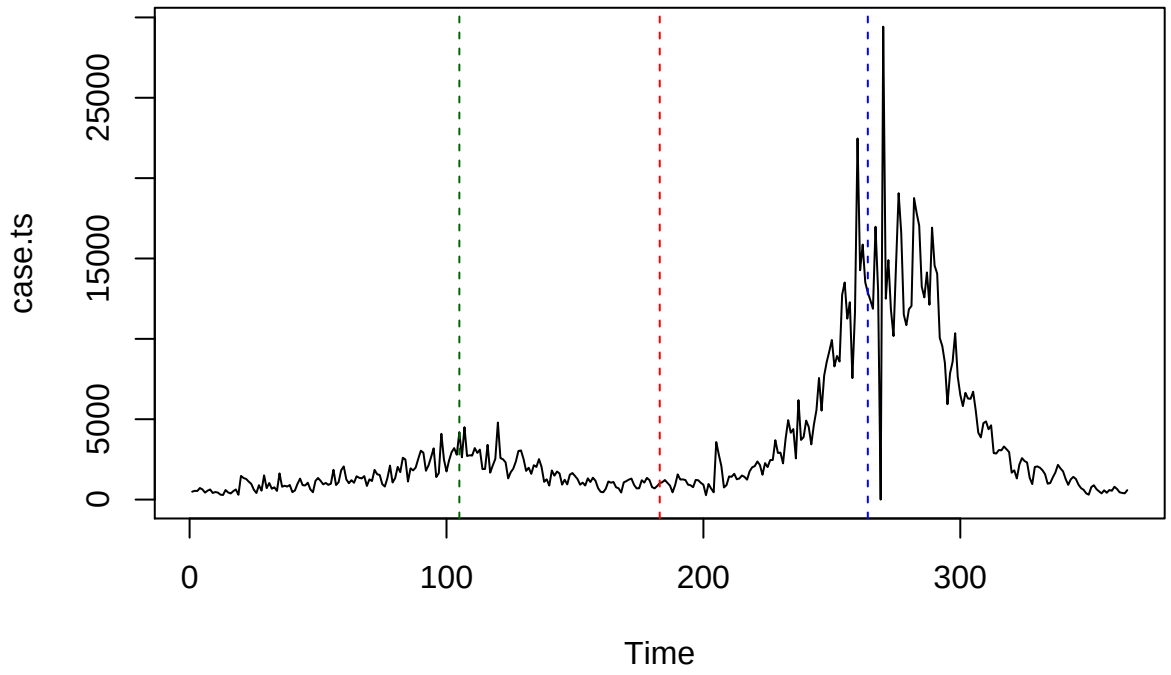
Time Series Final Project: Exploratory Data Analysis

Annie Cohen and Lian Morales

4/19/2021

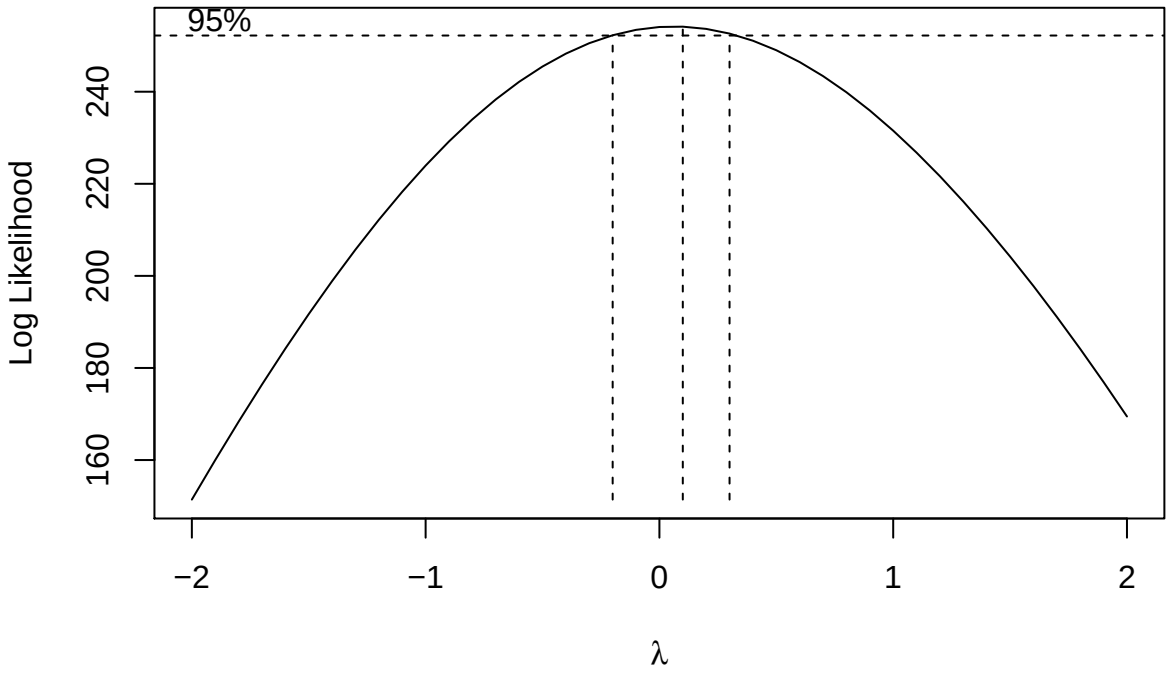
In our project, we are taking data from the LA Times Github repository of COVID cases in the LA county (https://github.com/datadesk/california-coronavirus-data). We will be splitting the data up into four sections and completing our analyses for each subset of the data. In this code, we will complete a BoxCox test to measure significance, data transformation, and confidence intervals to help us understand the data and also how to best analyze it. We will also perform acf and pacf to determine seasonality and check for major outliers and trends in the data.

```
# pulling the data from the Los Angeles County GitHub
casedata <- read.csv(text = getURL("https://raw.githubusercontent.com/datadesk/california-coronavirus-da
  filter(county == "Los Angeles") %>%
  mutate(date = date(date), month = month(date)) %>%
  map_df(rev) %>%
  filter(!is.na(new_confirmed_cases) & between(date, date("2020-04-01"),date("2021-03-31")))

# creating the time series
case.ts <- ts(casedata$new_confirmed_cases, start = 1,  frequency = 1)

plot(case.ts)
abline(v = 105, lty = 2, col = "darkgreen")
abline(v = 183, lty = 2, col = "red")
abline(v = 264, lty = 2, col = "blue")
```




```
# averaging dec 25th and 26th
case.ts[269] <- 14711
case.ts[270] <- 14712

## april 1 - sep 30
case.ts.1 <- ts(case.ts[1:183], start = 1, frequency = 1)

# april 1 - jul 14
case.ts.1.1 <- ts(case.ts[1:105], start = 1, frequency = 1)
# jul 15 - sep 30
case.ts.1.2 <- ts(case.ts[106:183], start = 1, frequency = 1)

## oct 1 - mar 31
case.ts.2 <- ts(case.ts[184:365], start = 1, frequency = 1)

# oct 1 - dec 20
case.ts.2.1 <- ts(case.ts[184:264], start = 1, frequency = 1)
# dec 21 - mar 31
case.ts.2.2 <- ts(case.ts[265:365], start = 1, frequency = 1)
```

```
## april 1 - july 14
bxcx.1.1 <- BoxCox.ar(case.ts.1.1) # error "possible convergence problem: optim gave code = 1"
```



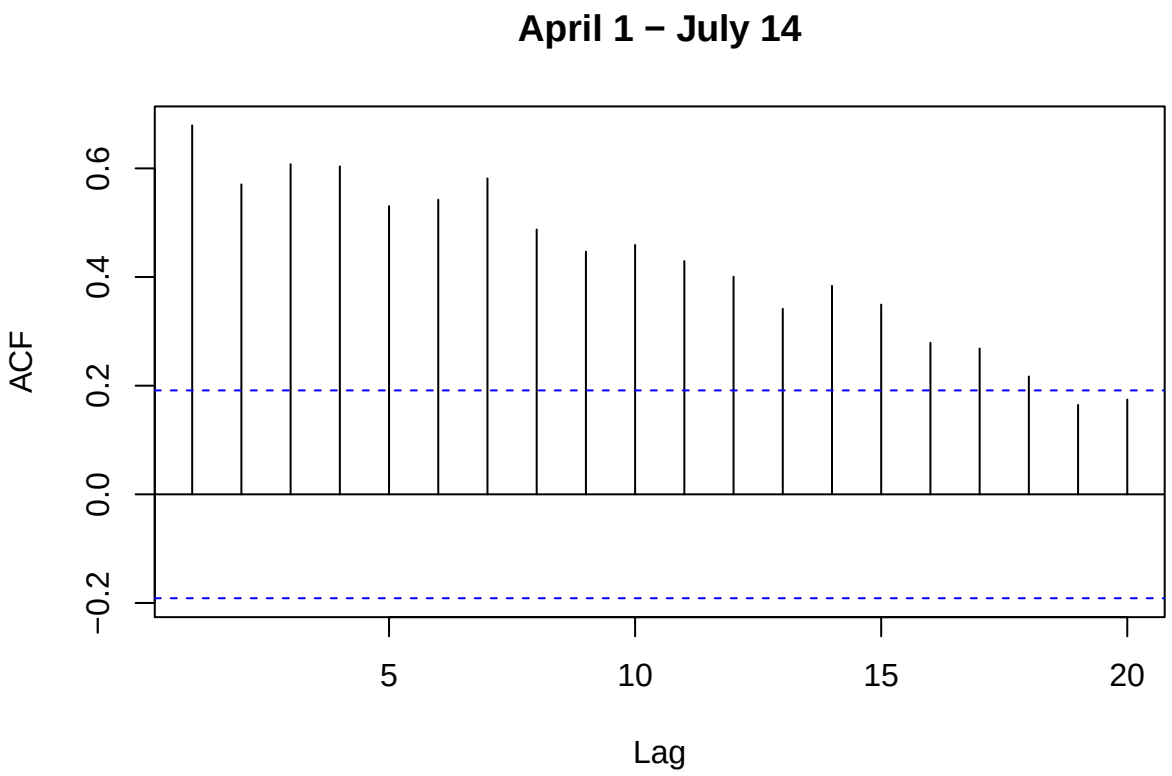

```
bxcx.1.1$mle
```

```
## [1] 0.1
```

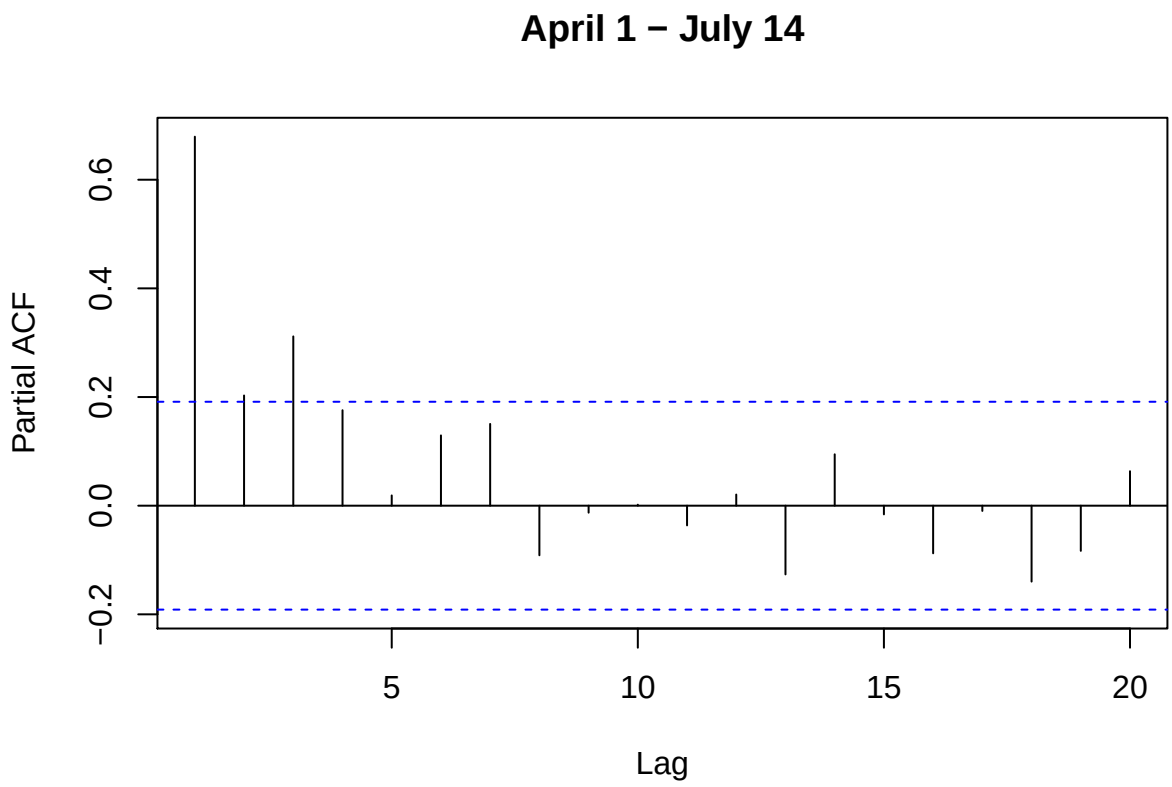
```
bxcx.1.1$ci
```

```
## [1] -0.2  0.3
```

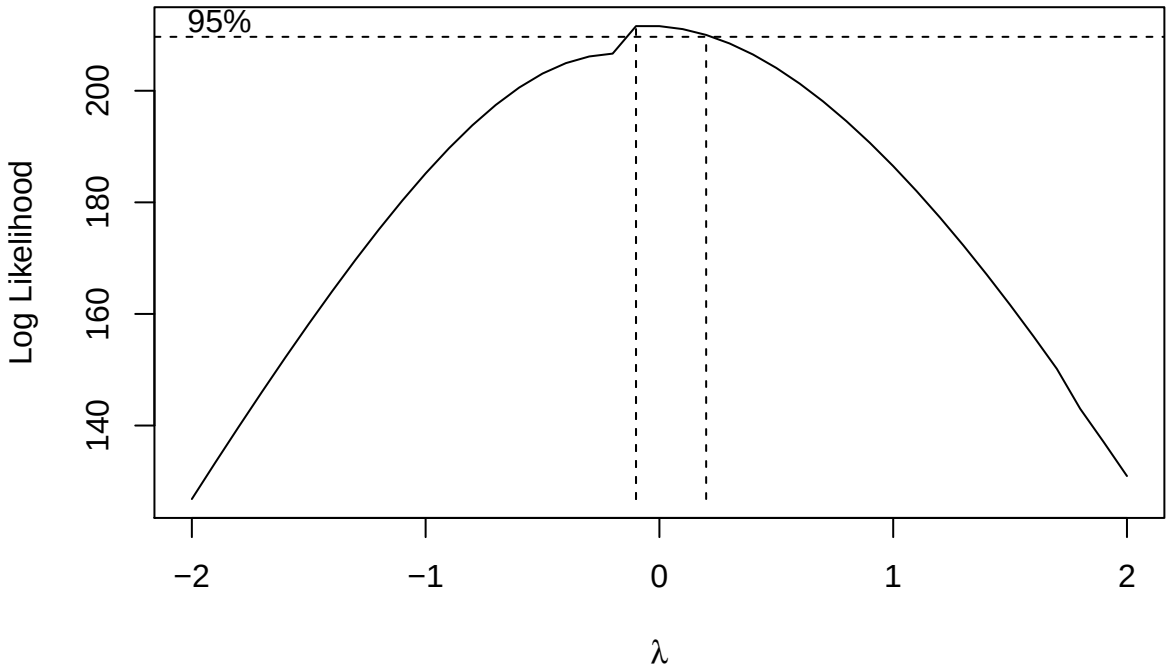
```
acf(case.ts.1.1, main = "April 1 - July 14")
```




```
pacf(case.ts.1.1, main = "April 1 - July 14")
```



```
## july 15 - september 30
bxcx.1.2 <- BoxCox.ar(case.ts.1.2)
```



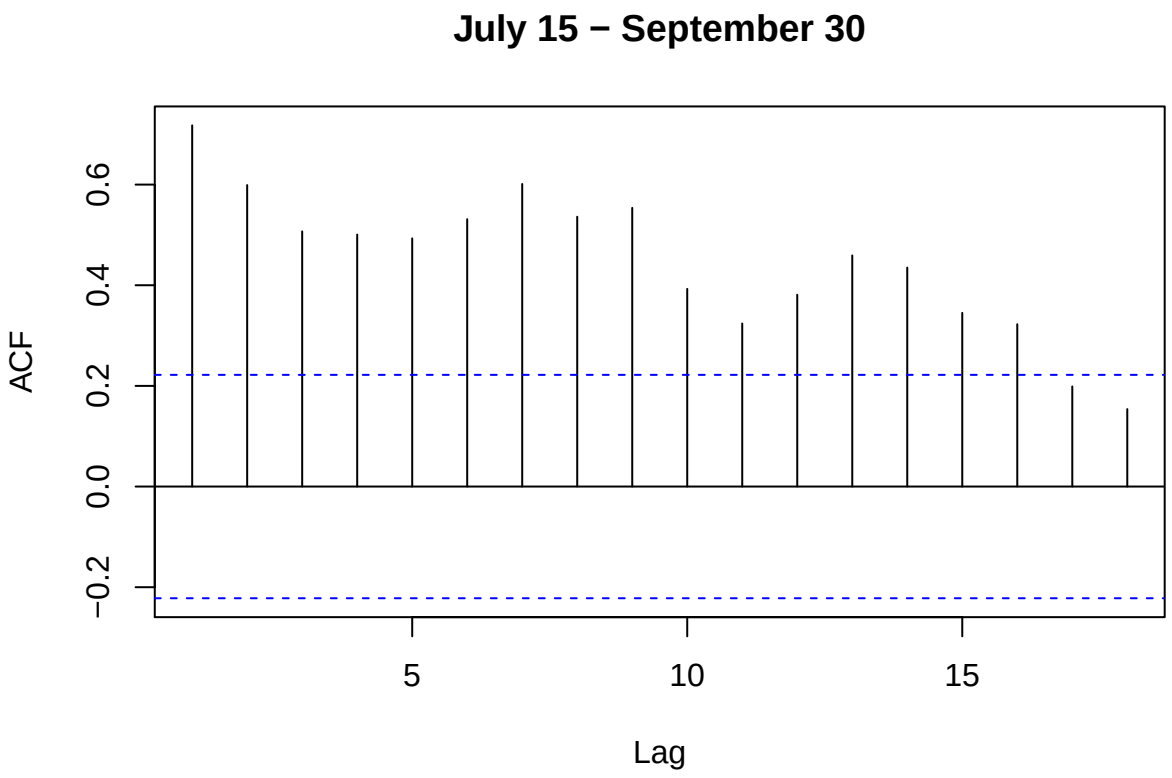

```
bxcx.1.2$mle
```

```
## [1] -0.1
```

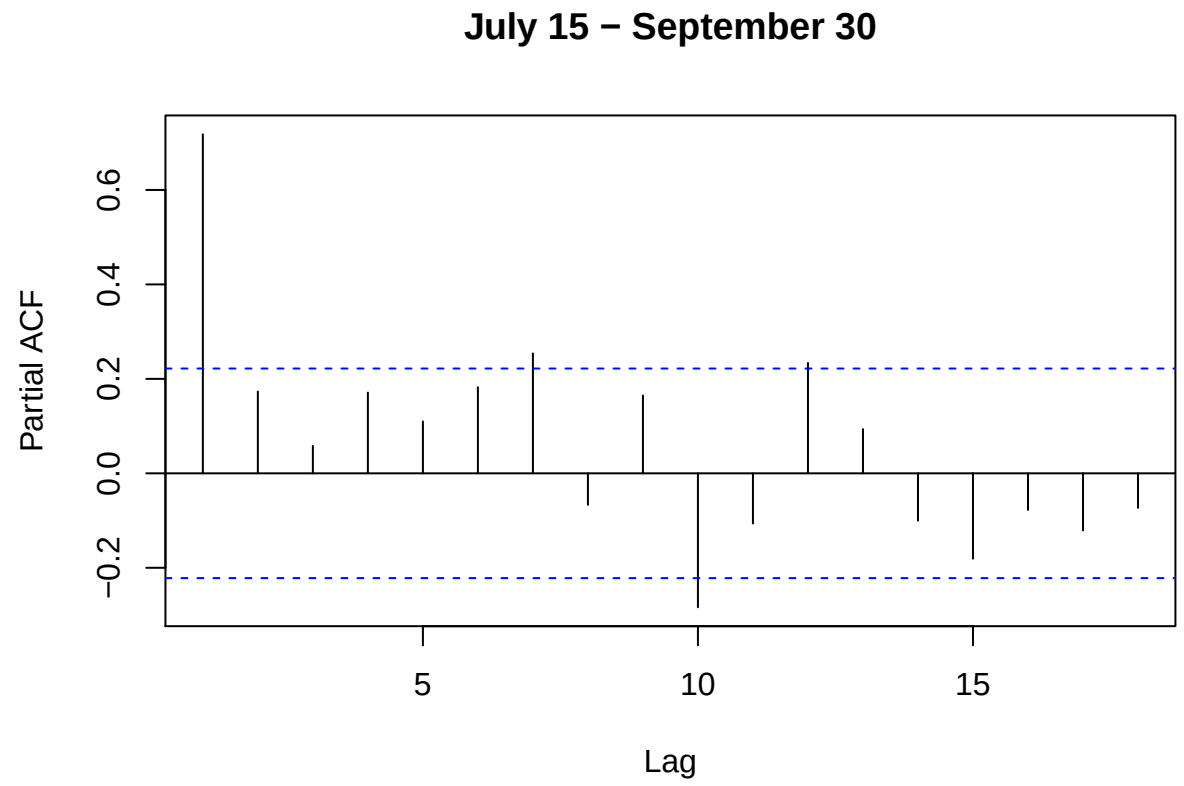
```
bxcx.1.2$ci
```

```
## [1] -0.1  0.2
```

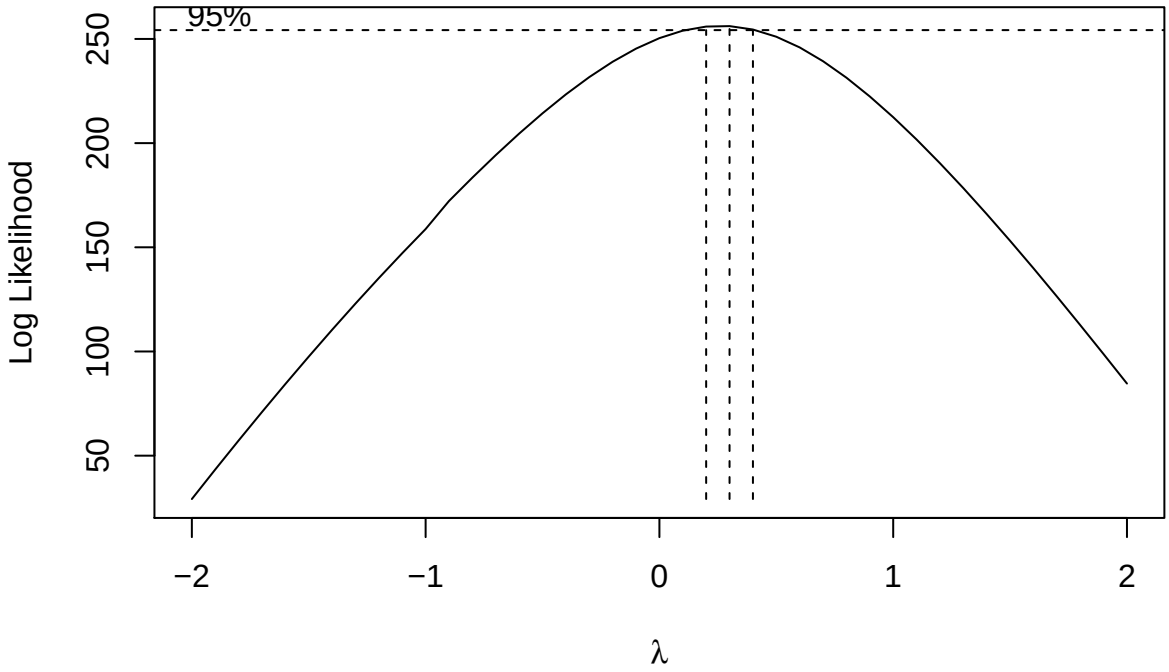
```
acf(case.ts.1.2, main = "July 15 - September 30")
```




```
pacf(case.ts.1.2, main = "July 15 - September 30")
```



```r
## october 1 - december 20
bxcx.2.1 <- BoxCox.ar(case.ts.2.1)
```



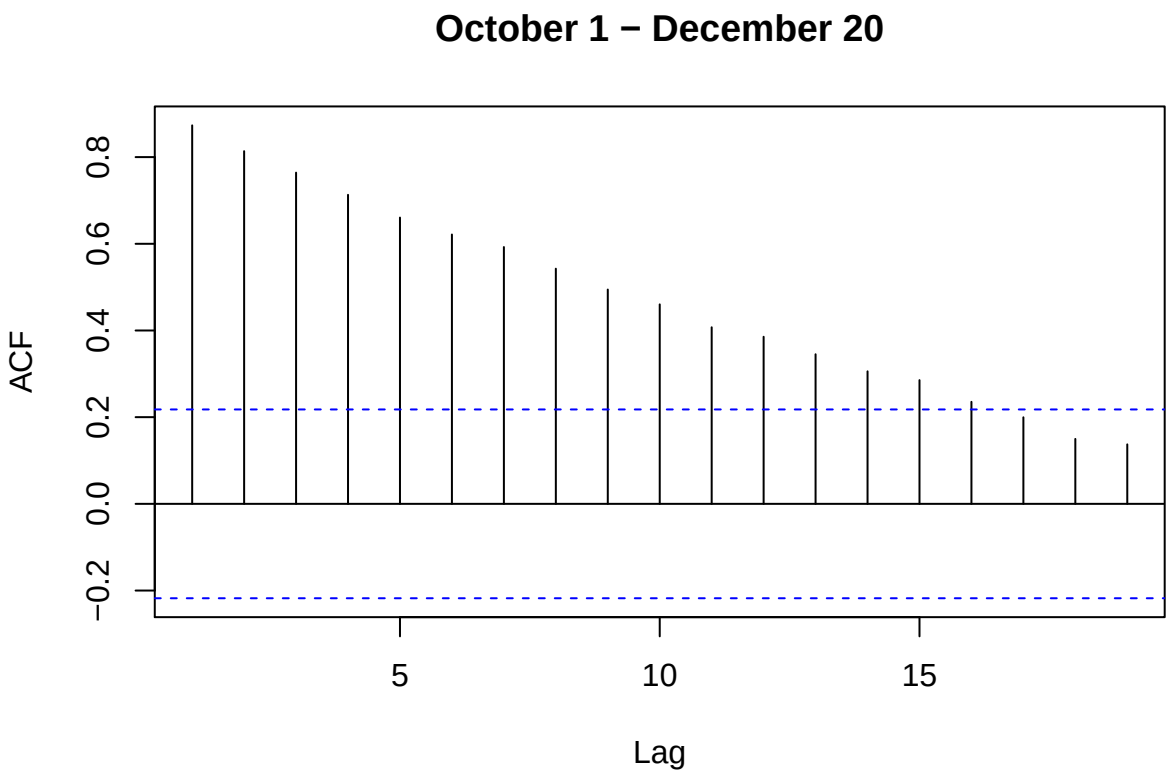

```r
bxcx.2.1$mle
```

```
## [1] 0.3
```

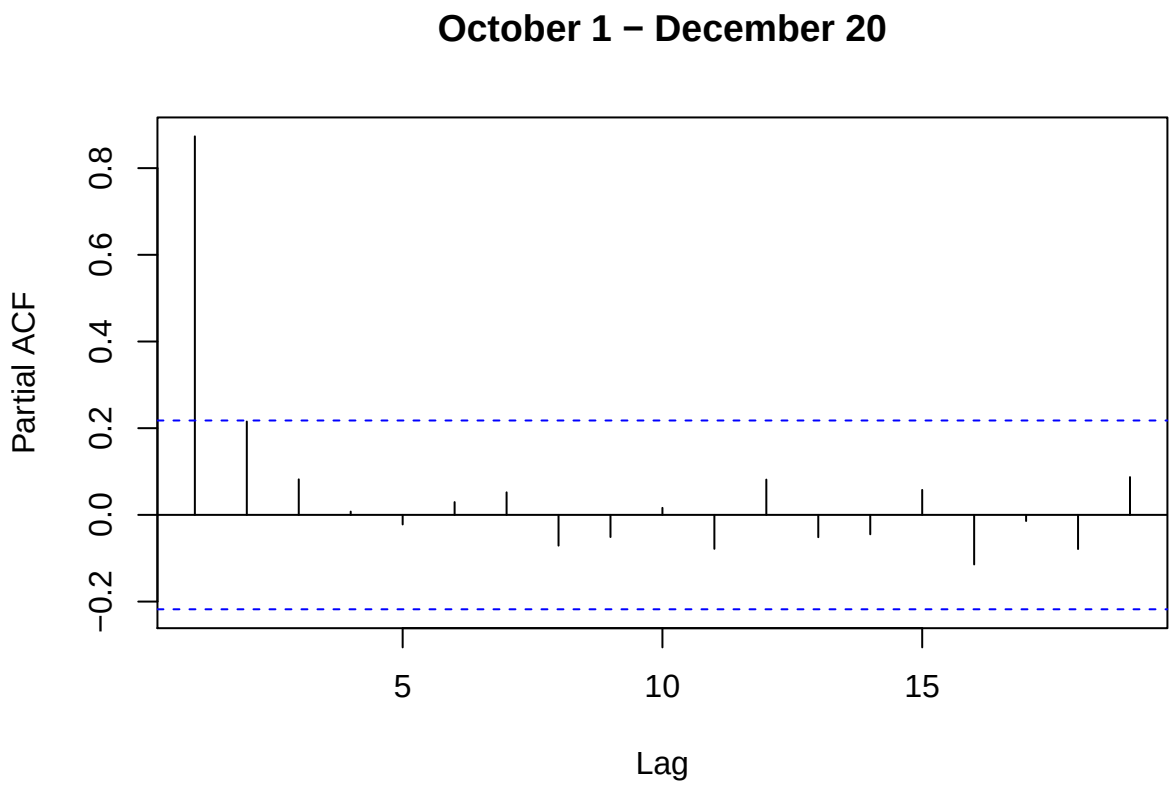
```r
bxcx.2.1$ci
```

```
## [1] 0.2 0.4
```

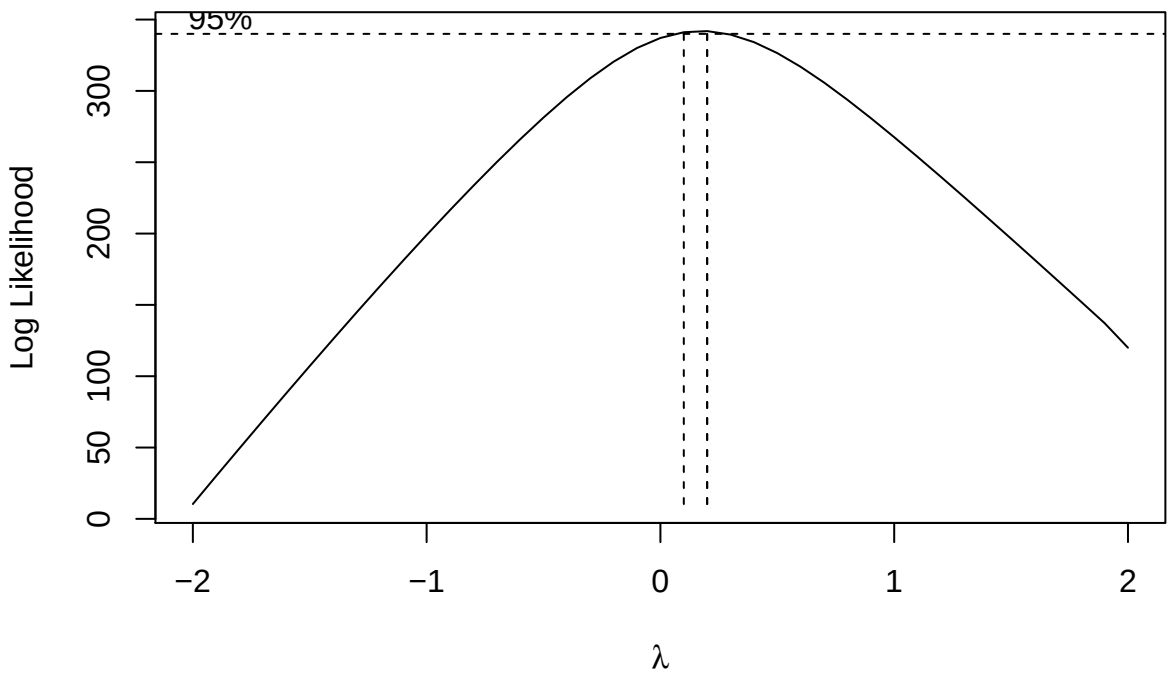
```r
acf(case.ts.2.1, main = "October 1 - December 20")
```




```
pacf(case.ts.2.1, main = "October 1 - December 20")
```



```
## december 21 - march 31
bxcx.2.2 <- BoxCox.ar(case.ts.2.2)
```



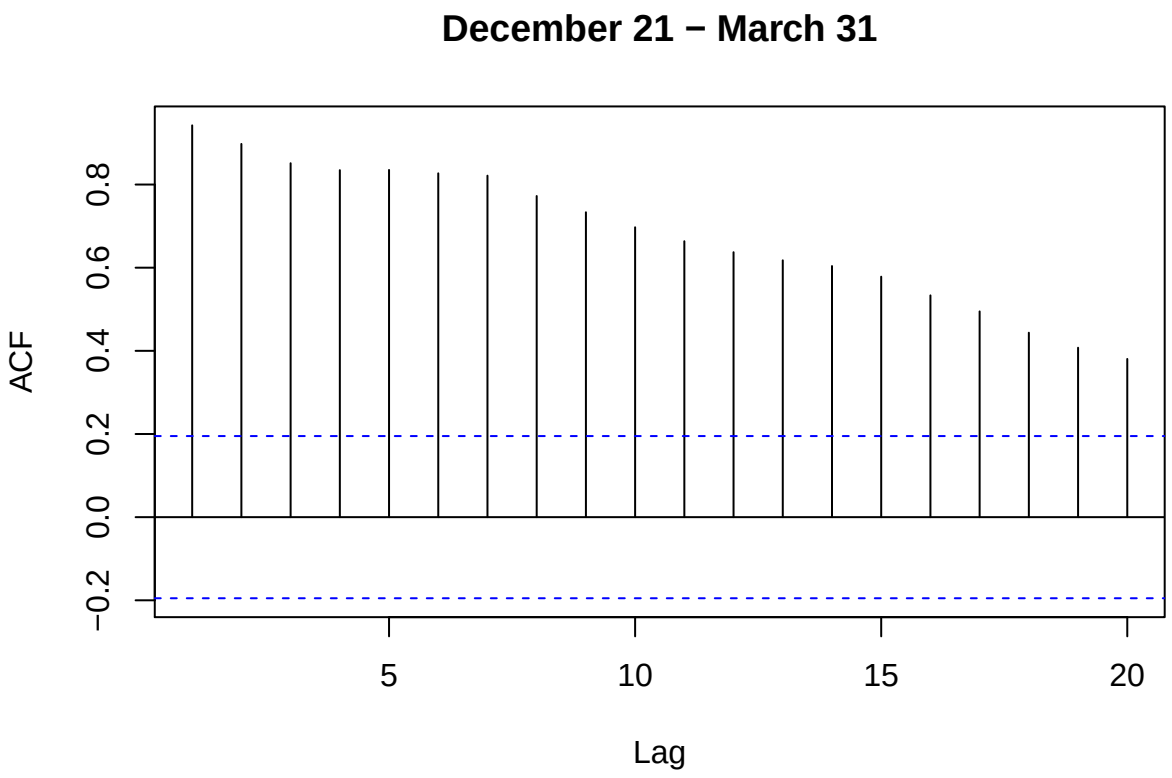

```
bxcx.2.2$mle
```

```
## [1] 0.2
```

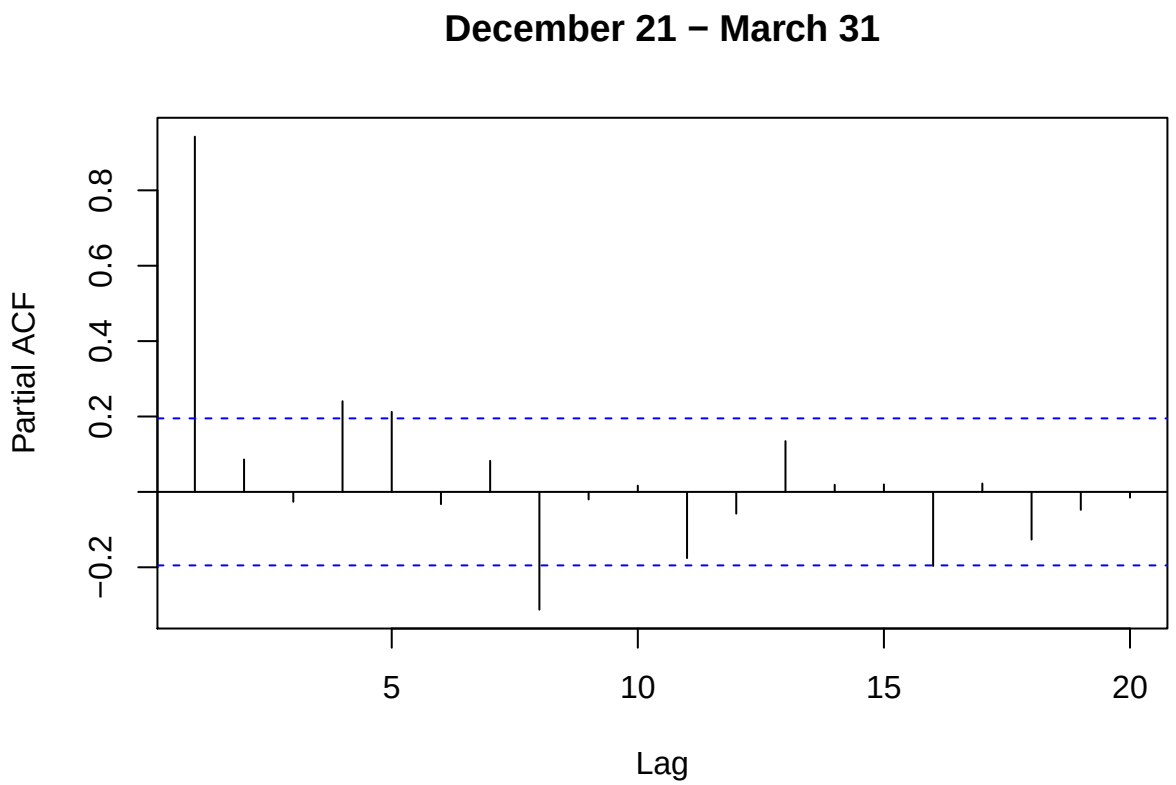
```
bxcx.2.2$ci
```

```
## [1] 0.1 0.2
```

```
acf(case.ts.2.2, main = "December 21 - March 31")
```




```
pacf(case.ts.2.2, main = "December 21 - March 31")
```



In the first BoxCox plot, April 1 - July 14, we notice that λ overlaps 0, with mle = 0.1. Therefore we will consider log transforming the data.

From July 15 to September 30, we will also consider a log transformation.

For October 1st to December 20, we will consider the mle transformation, 0.3.

And, for December 21 - march 31, the mle is 0.2, which we will transform our data to.

In each of the acf and pacf's of these subsets of the dataset, there does not seem to show white noise process, however in the pacf there seems to be a seasonal, weekly pattern with spikes around 7, 14 and 21. For the ongoing analysis, we will try to transform these sections of the data to best fit a time series, seasonal model.