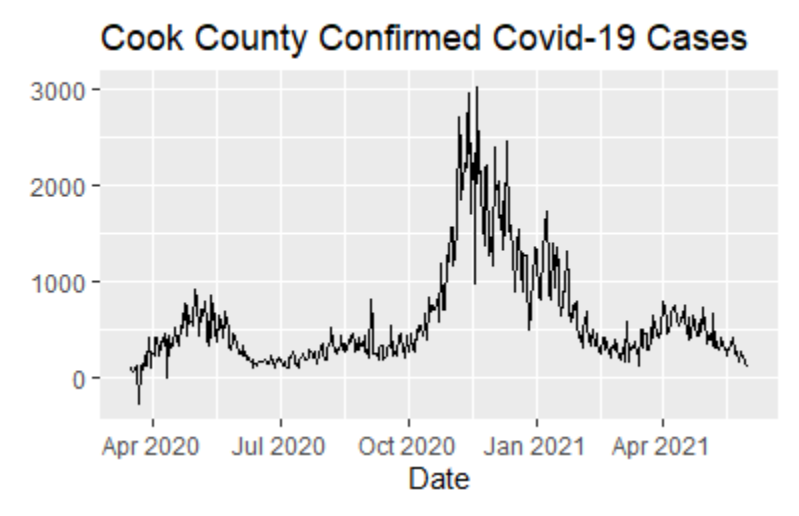
# Cook County Analysis

## Data Breakdown

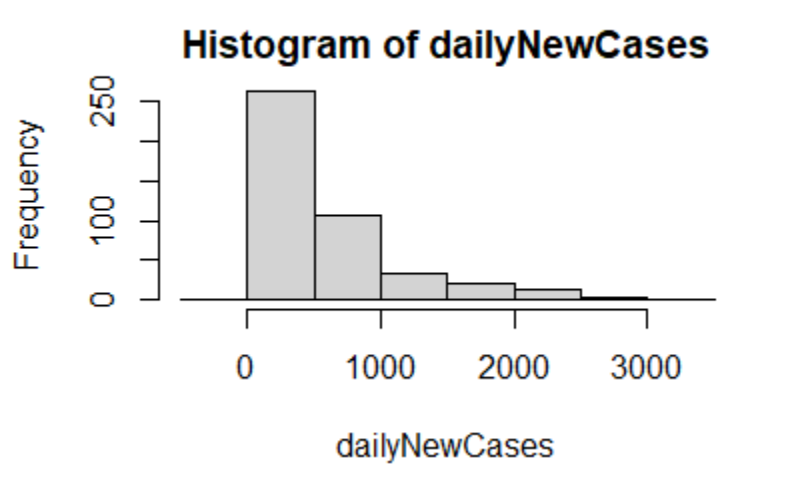
* Covid-19 Time Series Data from Cook County, Illinois
* Columns:
  + County Name
  + Total # Tested (Cumulative)
  + Confirmed Cases (Cumulative)
  + Deaths (By Day)
  + Date

## Time Series Plot



* Note: Data was cumulative so the daily new cases were calculated by subtracting previous data points.
  + Anomaly: The data contained negative values which represent correction of previous incorrectly high estimates. All these values were replacing with 0.

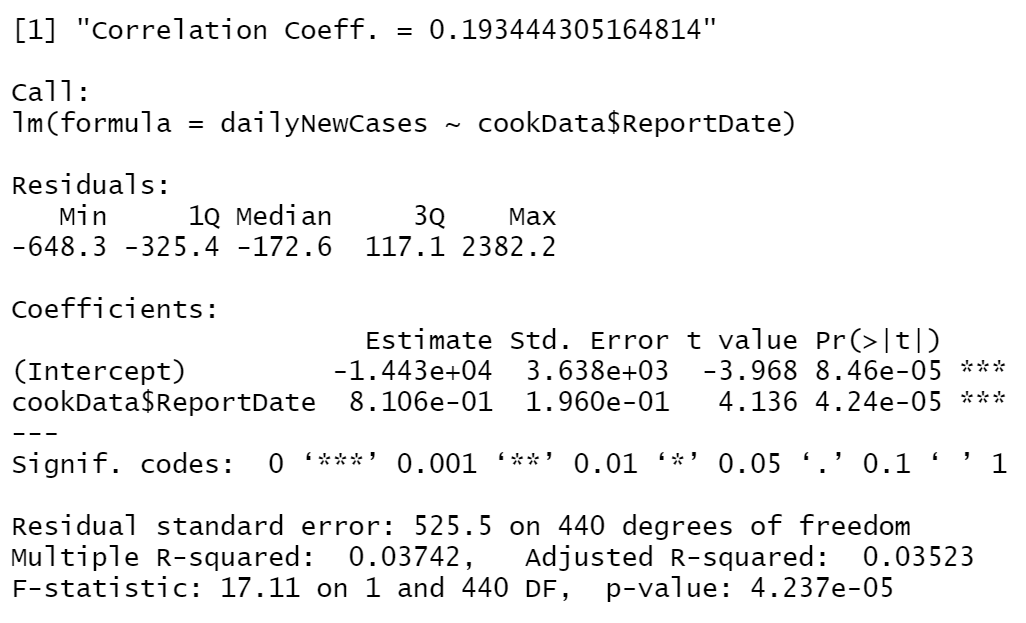
## Natural Data Skew

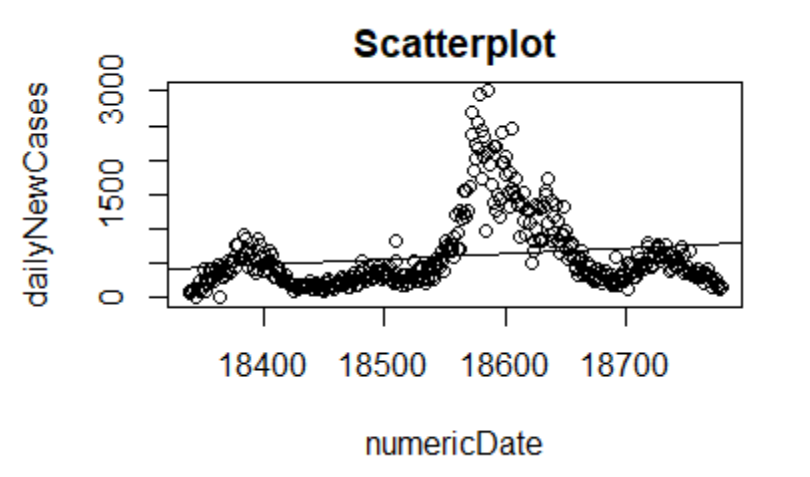
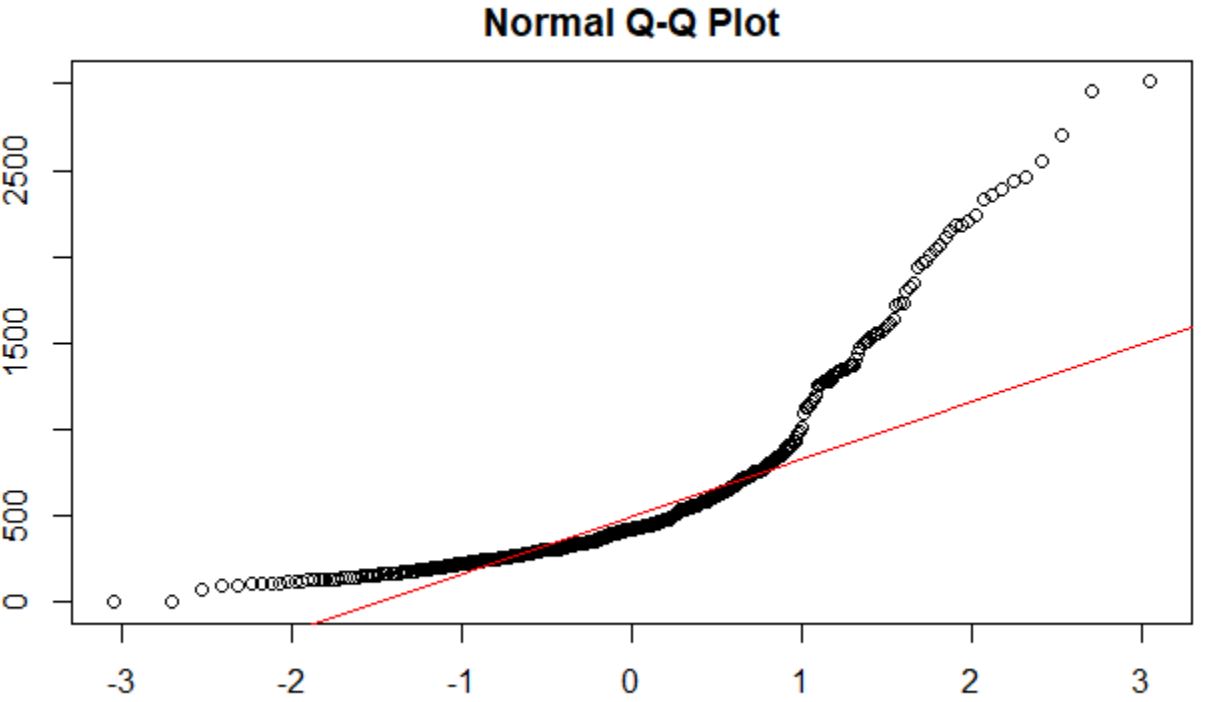


* As shown above the data is extremely positively skewed. Given that the relationship between x and y here is clearly not 1:1 we will attempt to transform the data to approximately normally distributed.

## Linear Regression on Natural Data

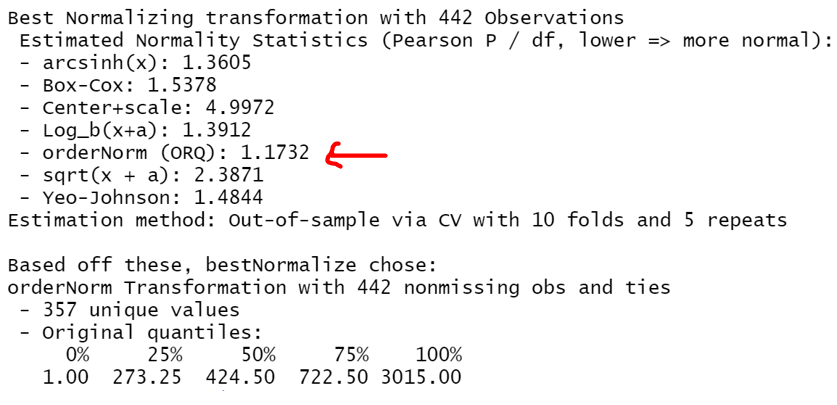
* The quartiles are not evenly distributed.
* T val is high so there is statistical significance?? (Obviously the longer the pandemic goes on the more people will be infected)
* R squared value is extremely low: Only explains 3.5% of data
* P value is smaller than 0.05 so there is statistical significance

 Regression on Data(as found)

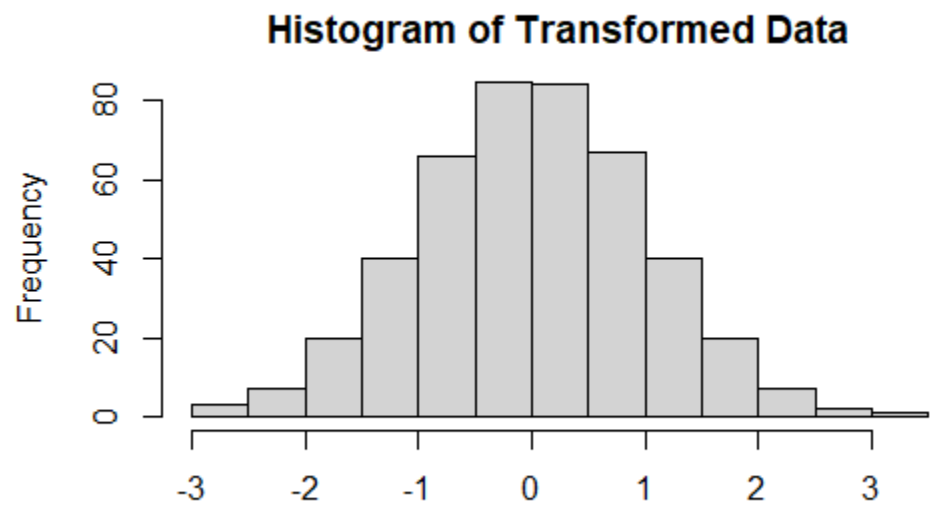
* Regression line doesn’t map on well
* 
* QQline doesn’t map on well. About half the data is off
* 

## Box Cox/Best Normalize

* Note: Estimate lambda for equation. Should be (-2,2). If 0 = use log10
* The correlation coefficient is abysmal so, we need to find lambda
  + Correlation Coeff. = 0.193444305164814
* I used the bestNormalize library to find a better lambda.



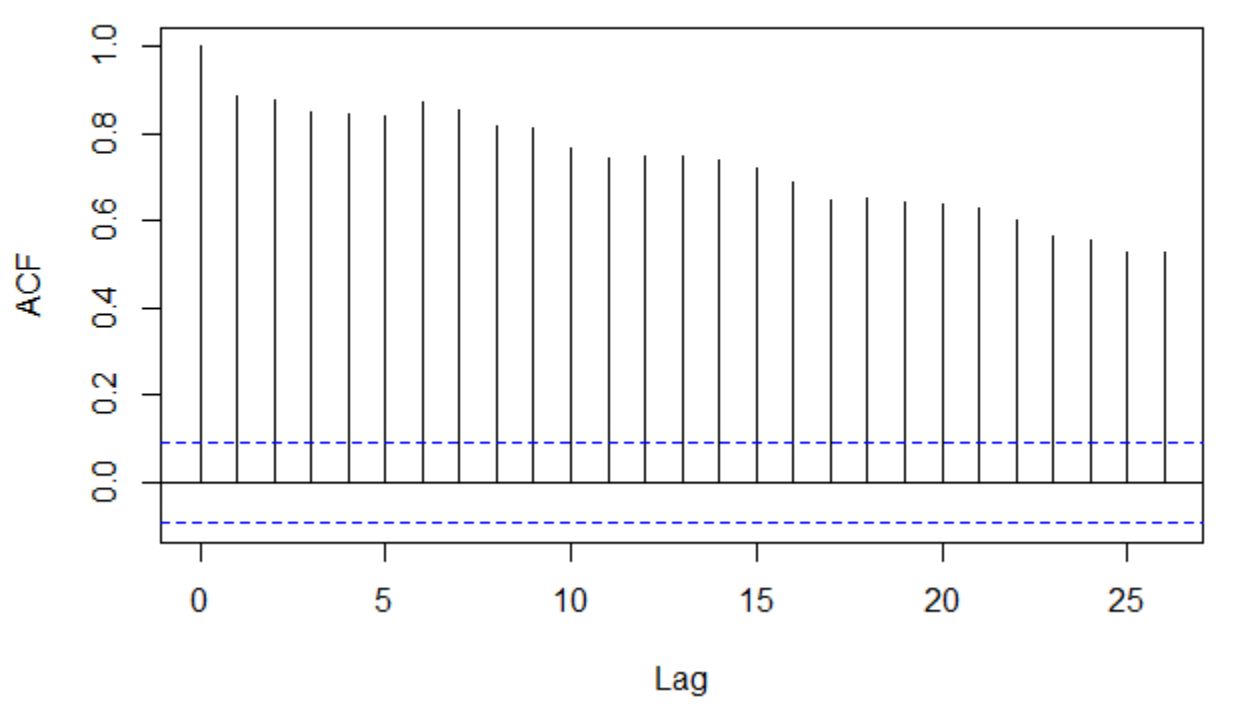
* That returns this gorgeous graph

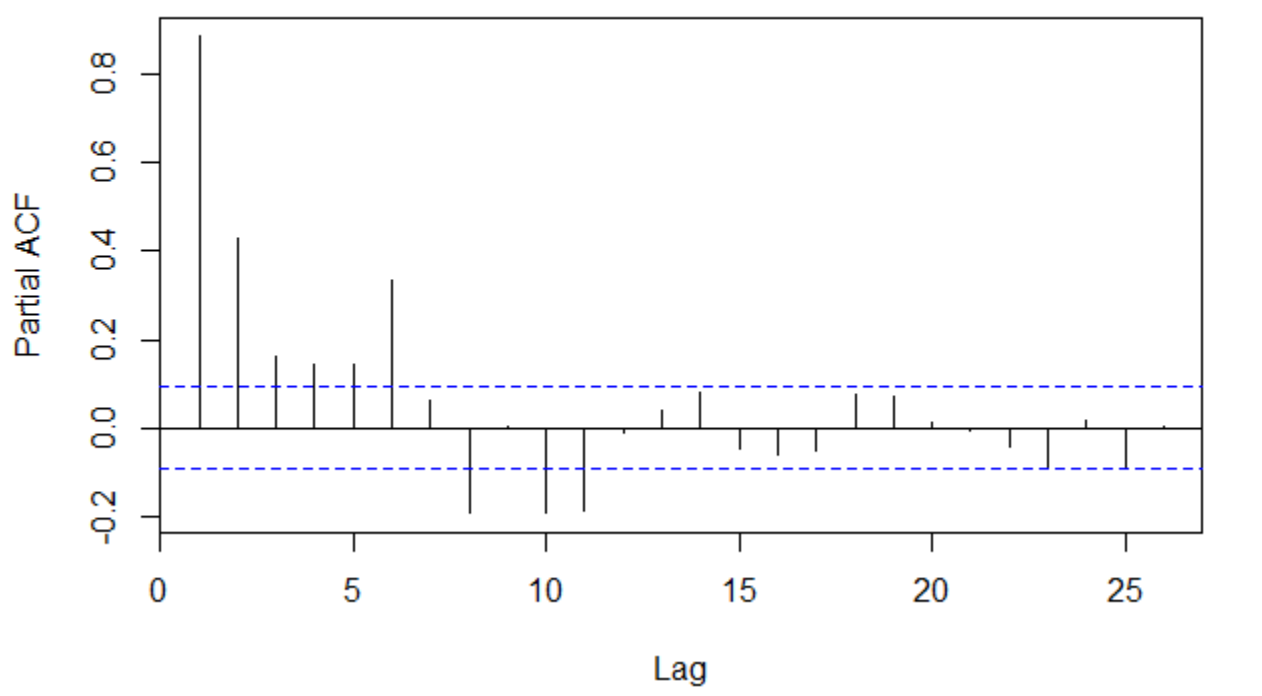


* How to find actual lamba???

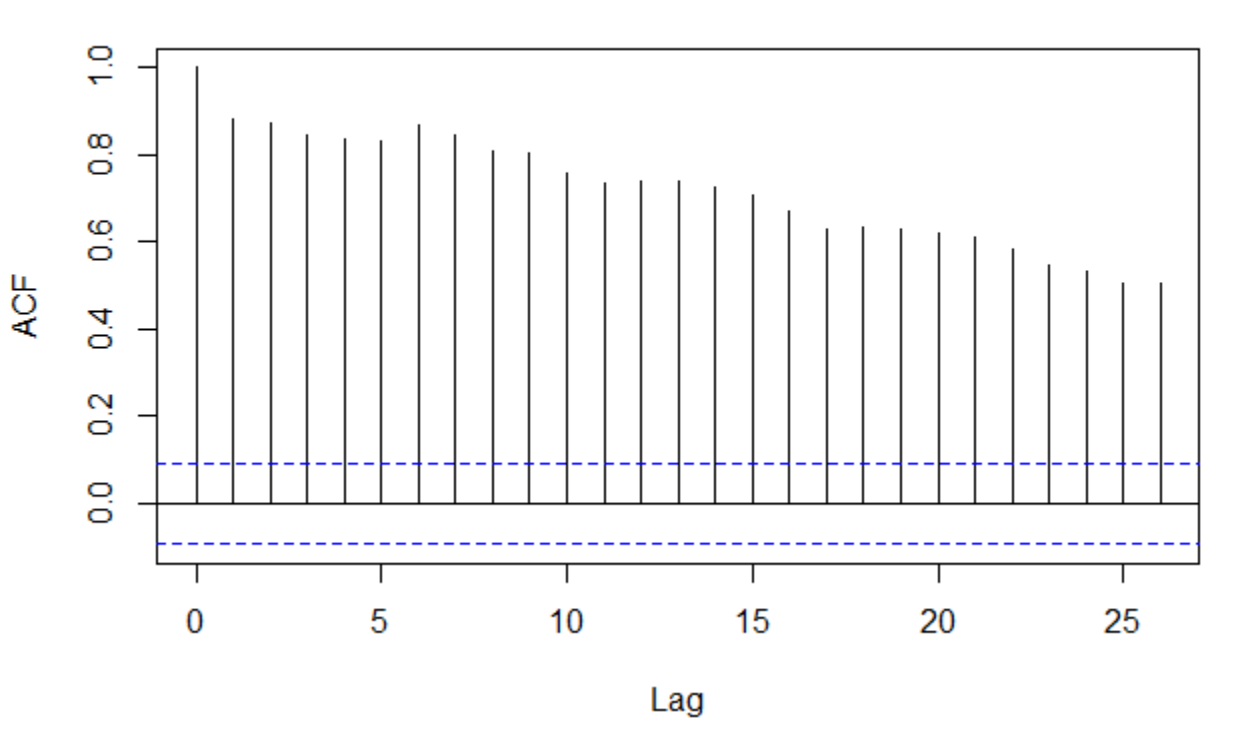
## ACF/PACF

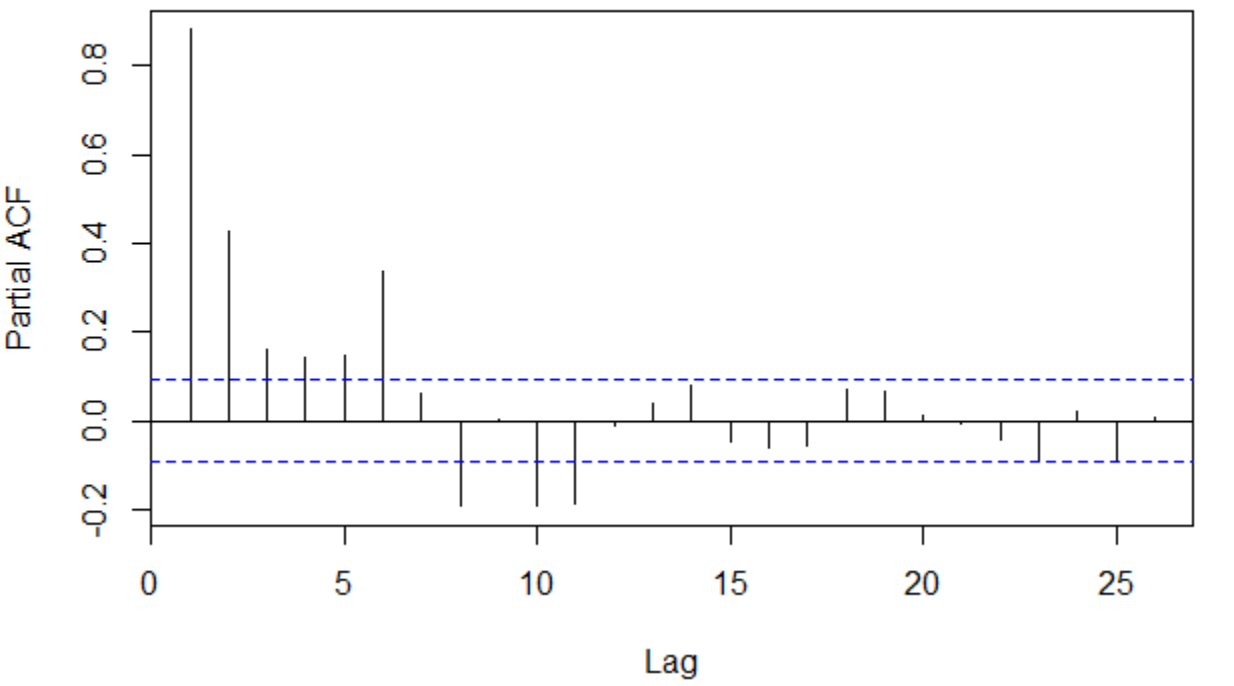
* ACF and PACF on natural data

ACF

PACF

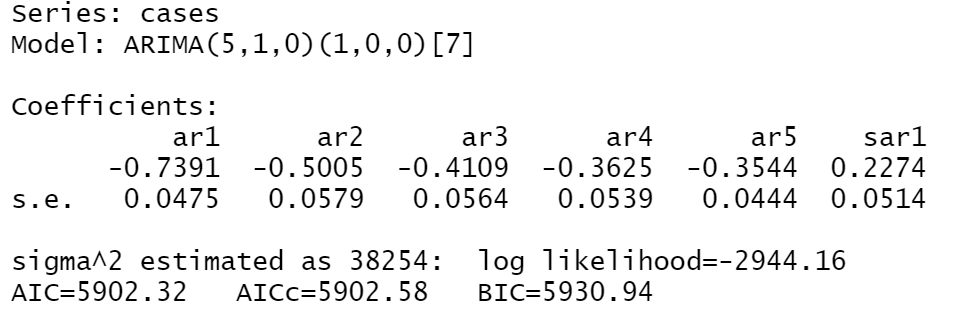
* ACF and PACF on new model

ACF

PACF

## Auto Regressive Model

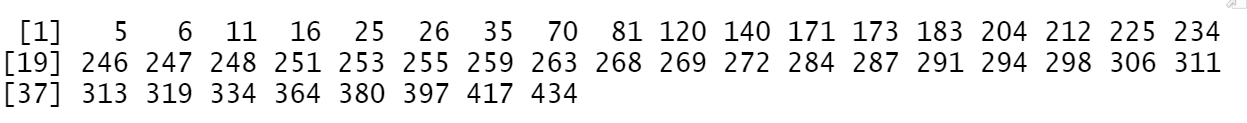
* https://nwfsc-timeseries.github.io/atsa-labs/sec-tslab-autoregressive-ar-models.html
* Auto regressive model with auto regressive of order 5 with seasonal auto regression of 1



## Changepoint

Changepoint mean.

* Penalty Value = 35



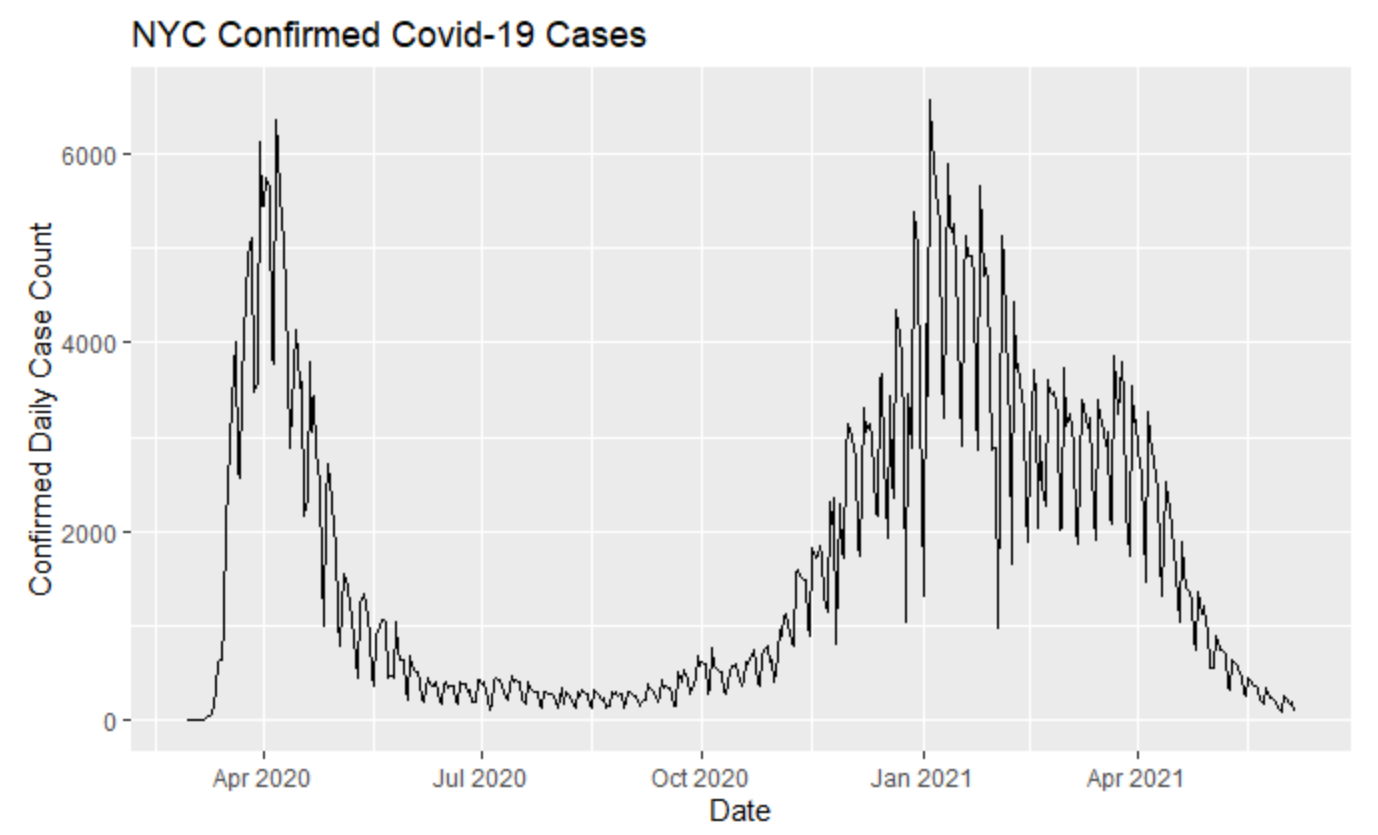
Spatial correlation

# NYC Analysis

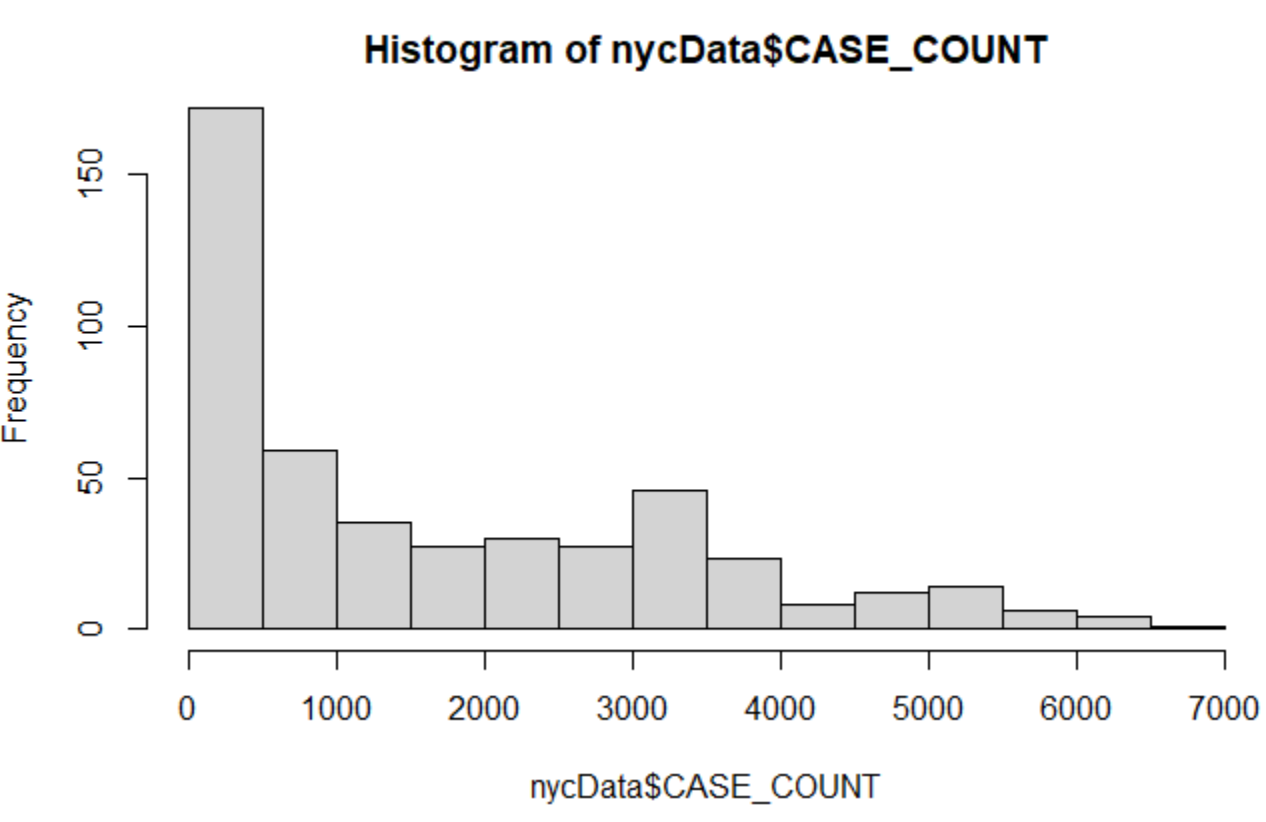
## Data Breakdown

* Covid-19 Time Series Data from Cook County, Illinois
* Columns(Relevant):
  + Date
  + Daily Positive Cases
  + Hospitalized
  + Death(By day)

## Time Series Plot



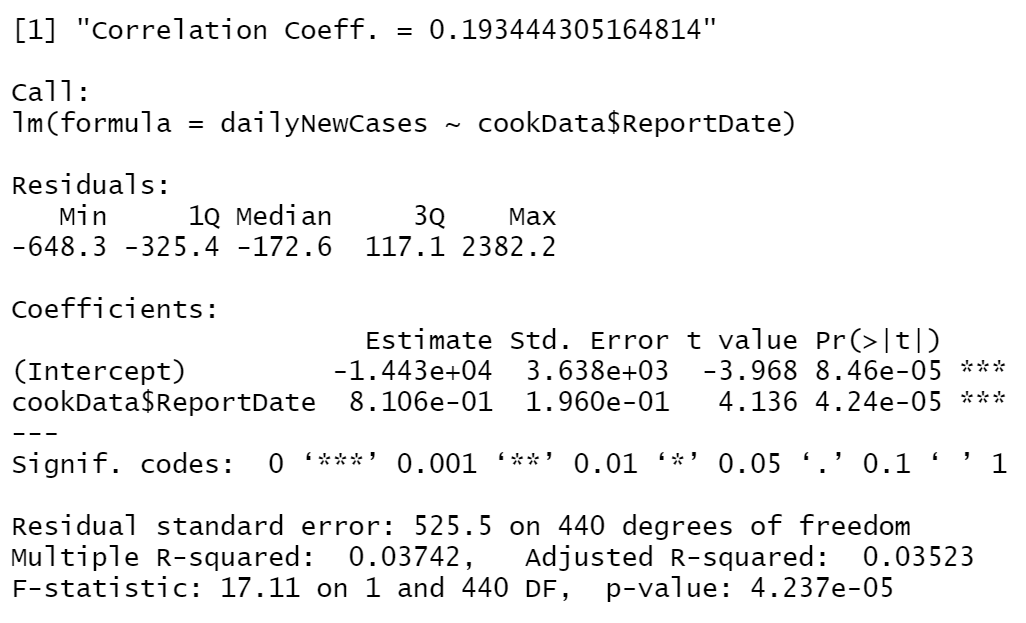
## Natural Data Skew



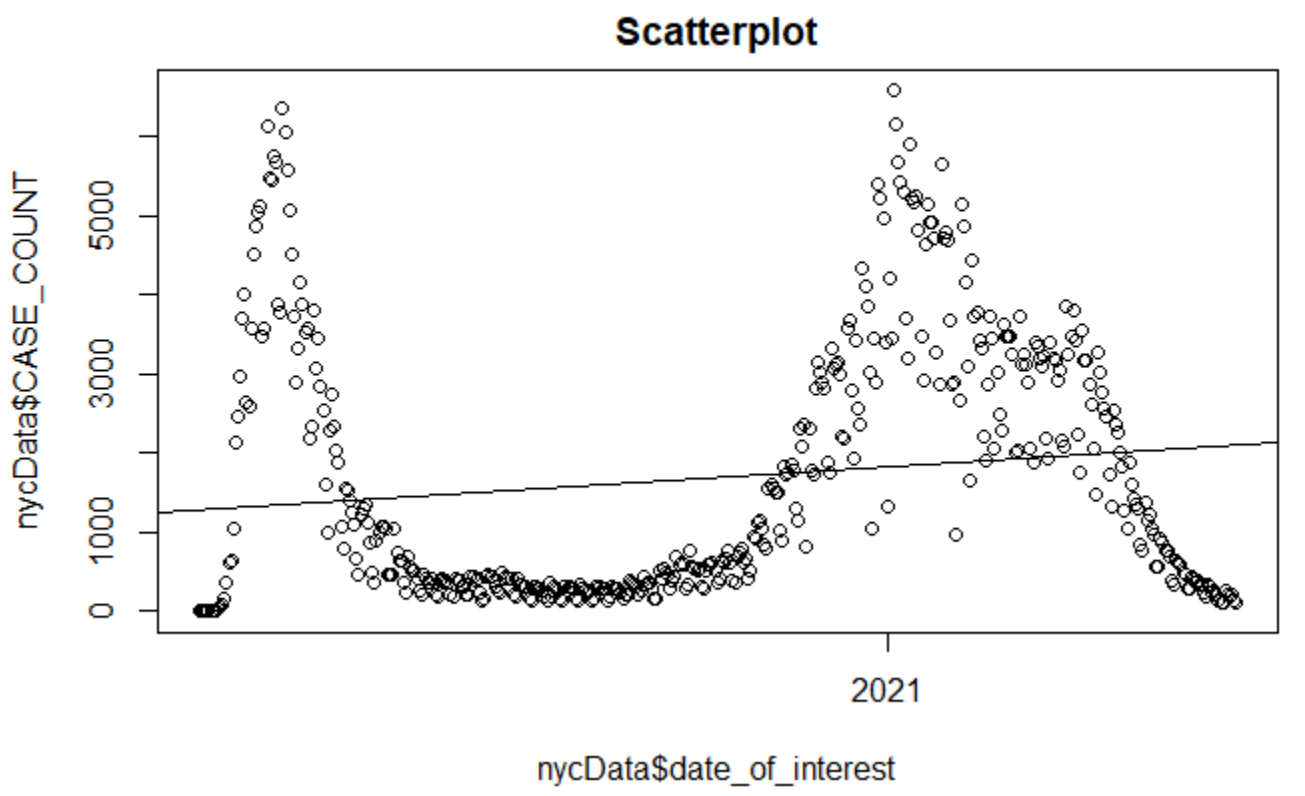
* As shown above the data is extremely positively skewed. Given that the relationship between x and y here is clearly not 1:1 we will attempt to transform the data to approximately normally distributed.

## Linear Regression on Natural Data

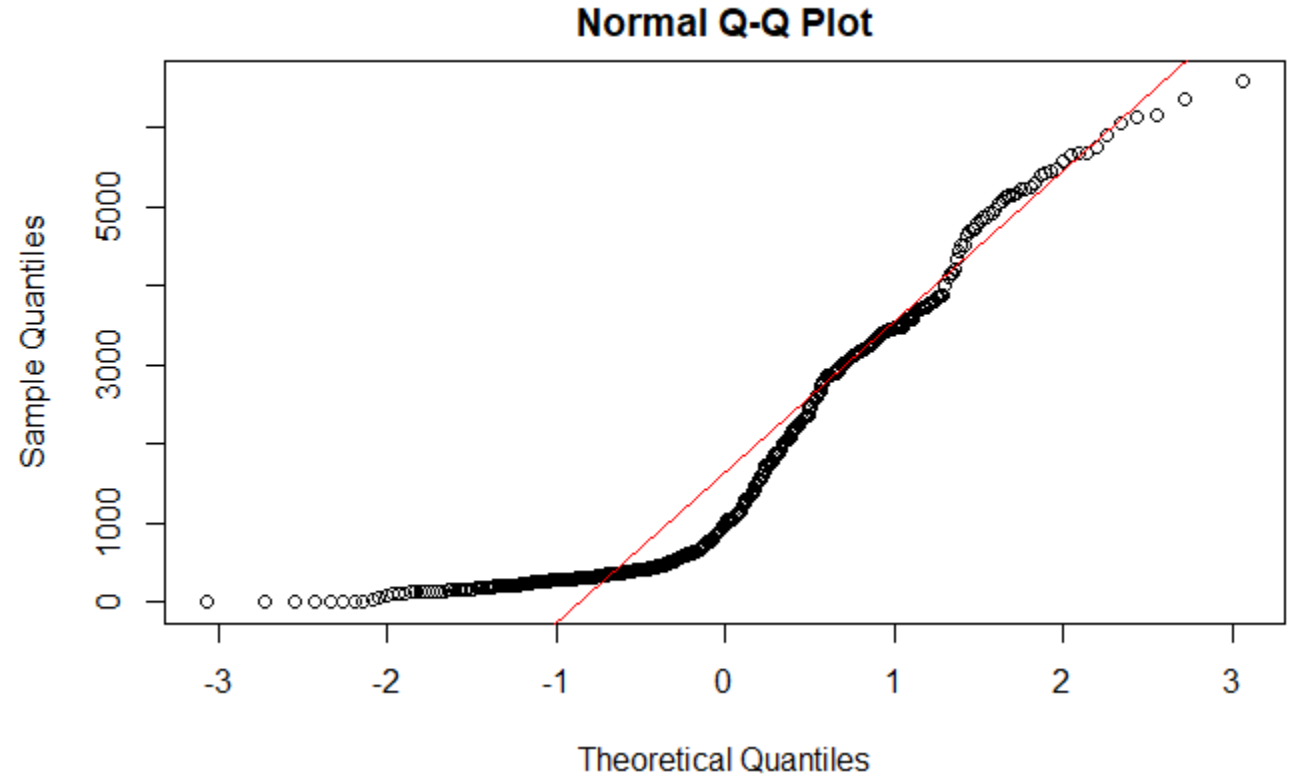
* Modeling the relationship between time and covid cases
* The quartiles are not evenly distributed.
* T val is high so there is statistical significance?? (Obviously the longer the pandemic goes on the more people will be infected)
* R squared value is extremely low: Only explains 3.5% of data
* P value is smaller than 0.05 so there is statistical significance

 Regression on Data (as found)

* Regression line doesn’t map on well

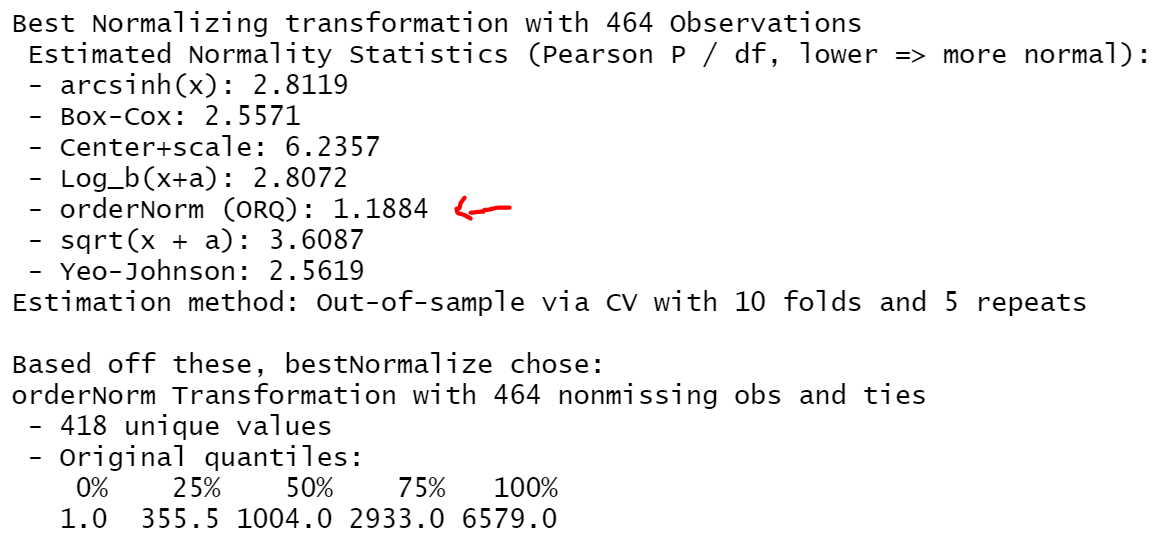


* QQline doesn’t map on well. About half the data is off

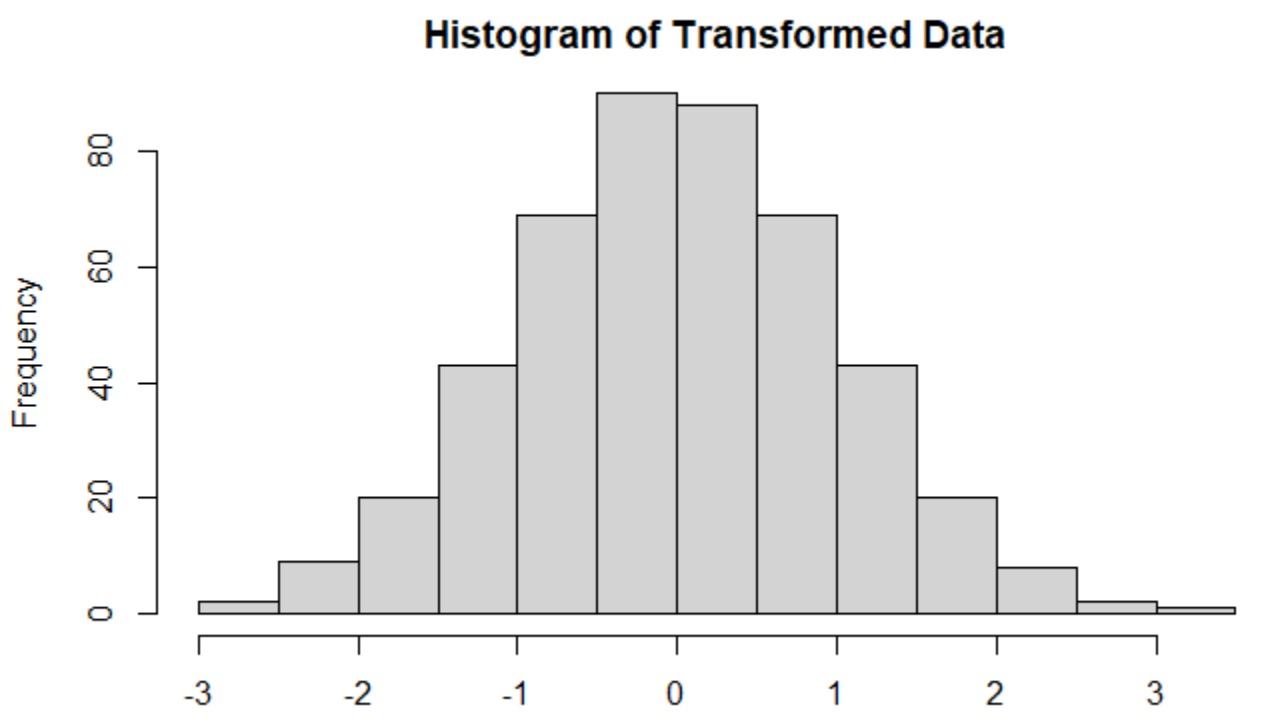


## Box Cox/Best Normalize

* Note: Estimate lambda for equation. Should be (-2,2). If 0 = use log10
* The correlation coefficient is abysmal so, we need to find lambda
  + Correlation Coeff. = 0.147361450148923
* I used the bestNormalize library to find a better lambda.



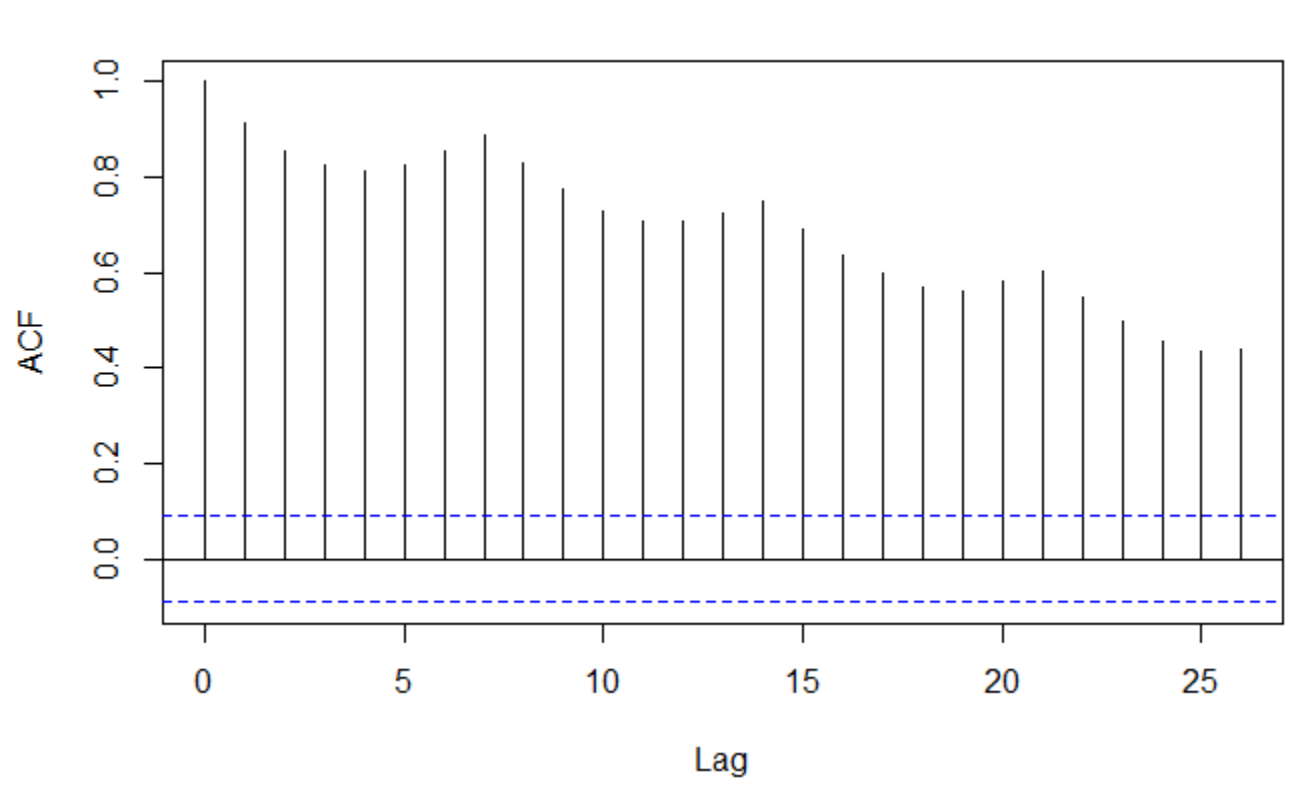
* That returns this gorgeous graph



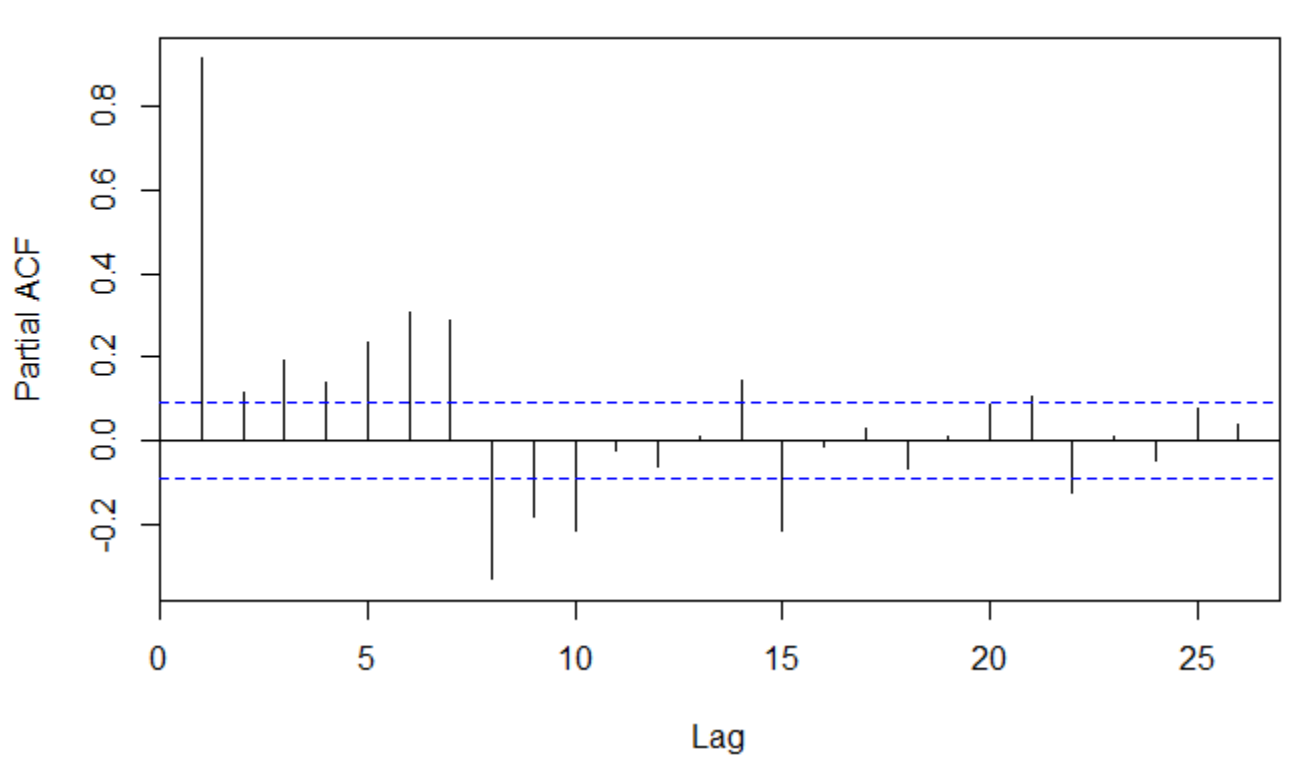
* Lamaba(0.2222222) doesn’t seem right???

## ACF/PACF

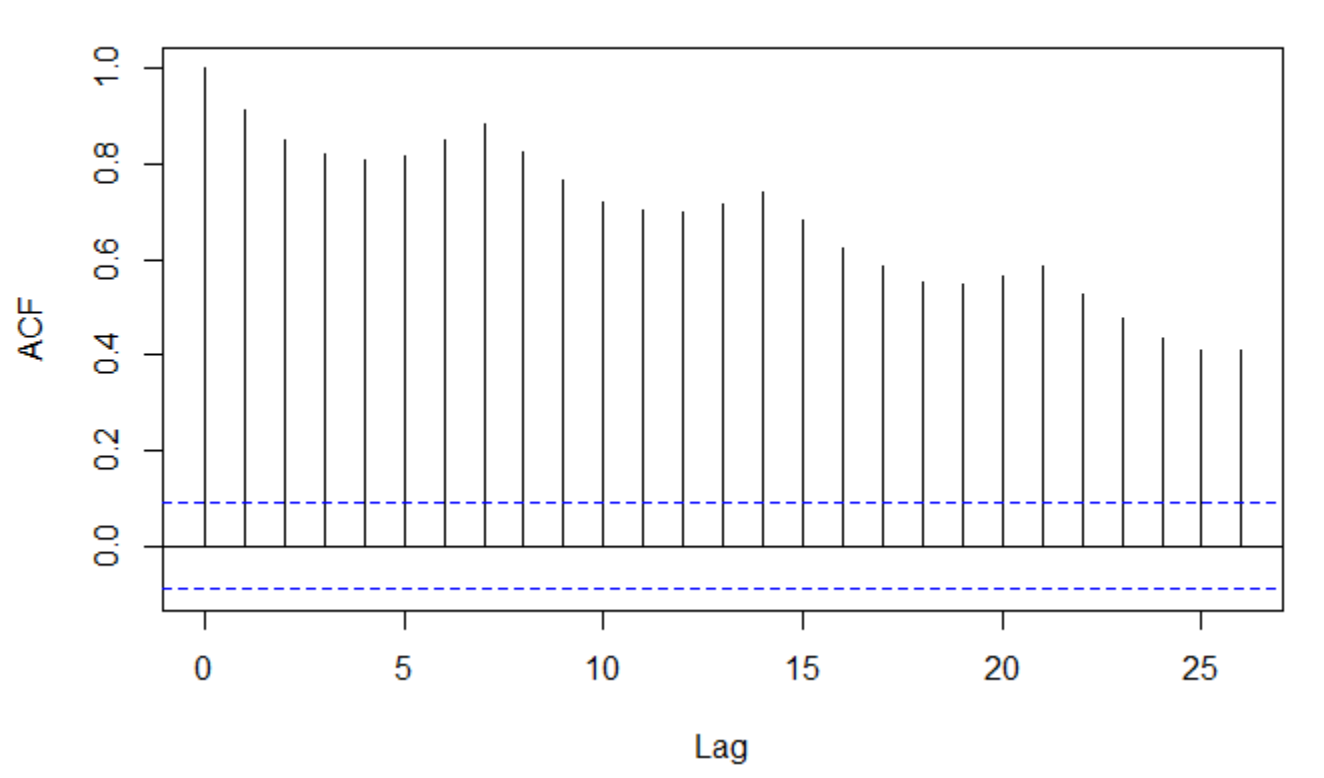
* ACF on natural data
* Data extremely similar even after 25 lags
* Since it is significantly above the blue line the data is statistically significant
* Is this AR(p) because amplitude of ACF trails off and PACF cuts off?

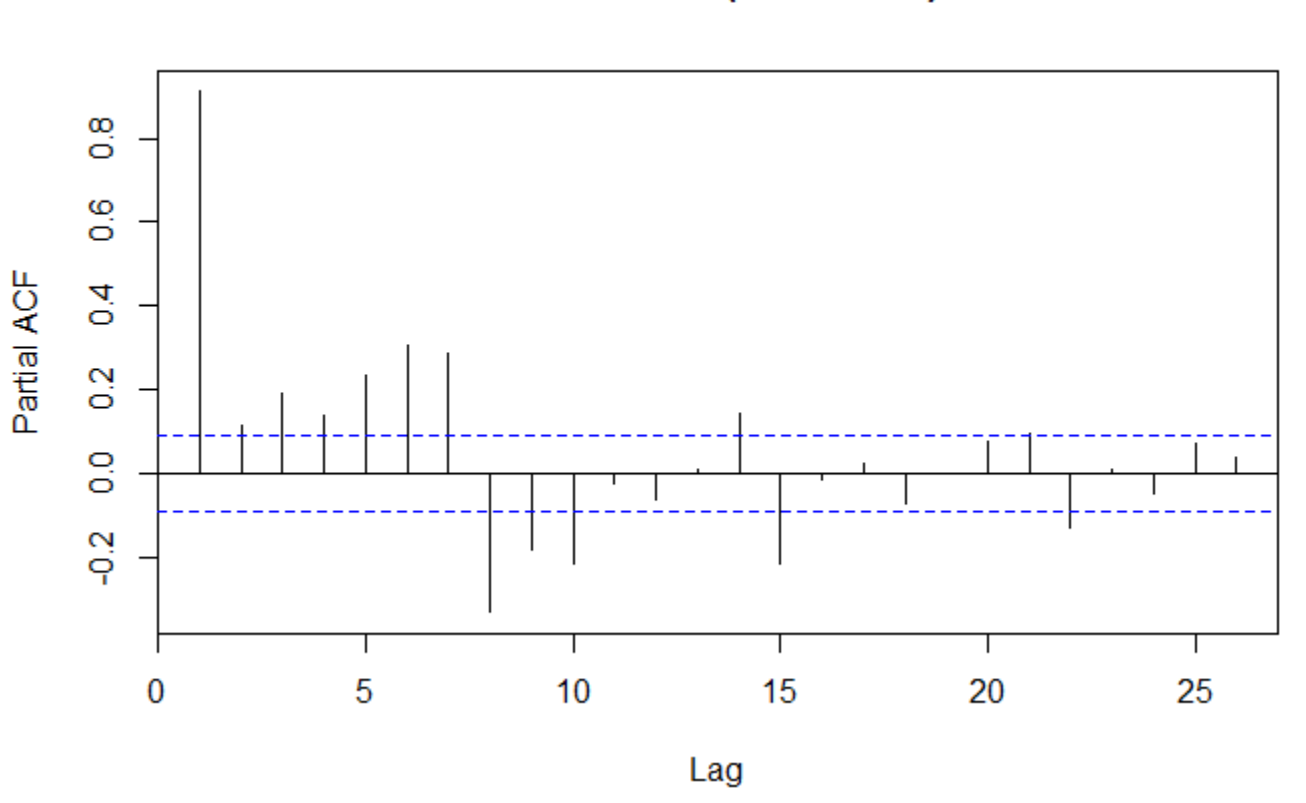
ACF

* PACF on natural data
* AR(2)

PACF

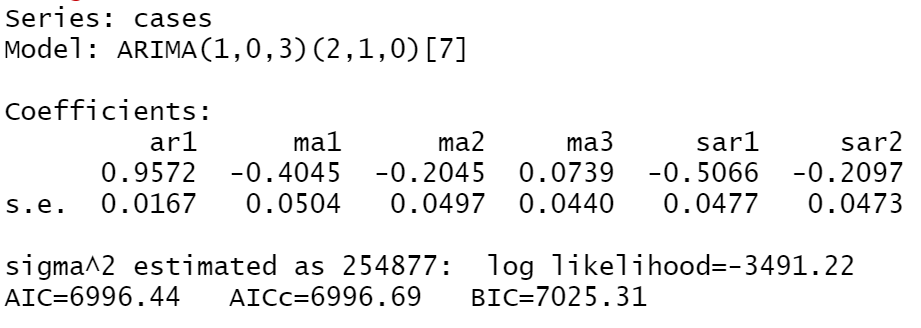
* ACF and PACF on new model
* Since it is significantly above the blue line the data is statistically significant
* Is this AR(p) because amplitude of ACF trails off and PACF cuts off?

ACF

PACF

## Auto Regressive Model

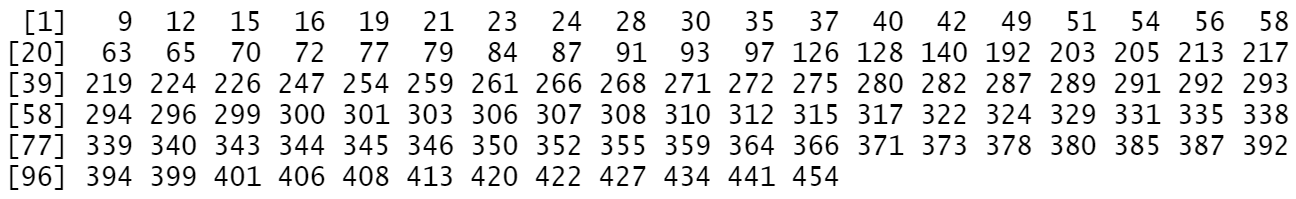
* <https://nwfsc-timeseries.github.io/atsa-labs/sec-tslab-autoregressive-ar-models.html>
* ARMA doesn’t work because there is seasonality involved in the data so ARIMA is better. Data isn’t super linear though
* Auto Regressive Integrated Moving Average
* (P(AR order), d(I), q(MA))
* Arima seasons
* Auto regressive model with auto regressive of order 1 with seasonal auto regression of 0
* How to make predictions???



## Changepoint

Changepoint mean.

* Penalty Value = 35



Spatial correlation