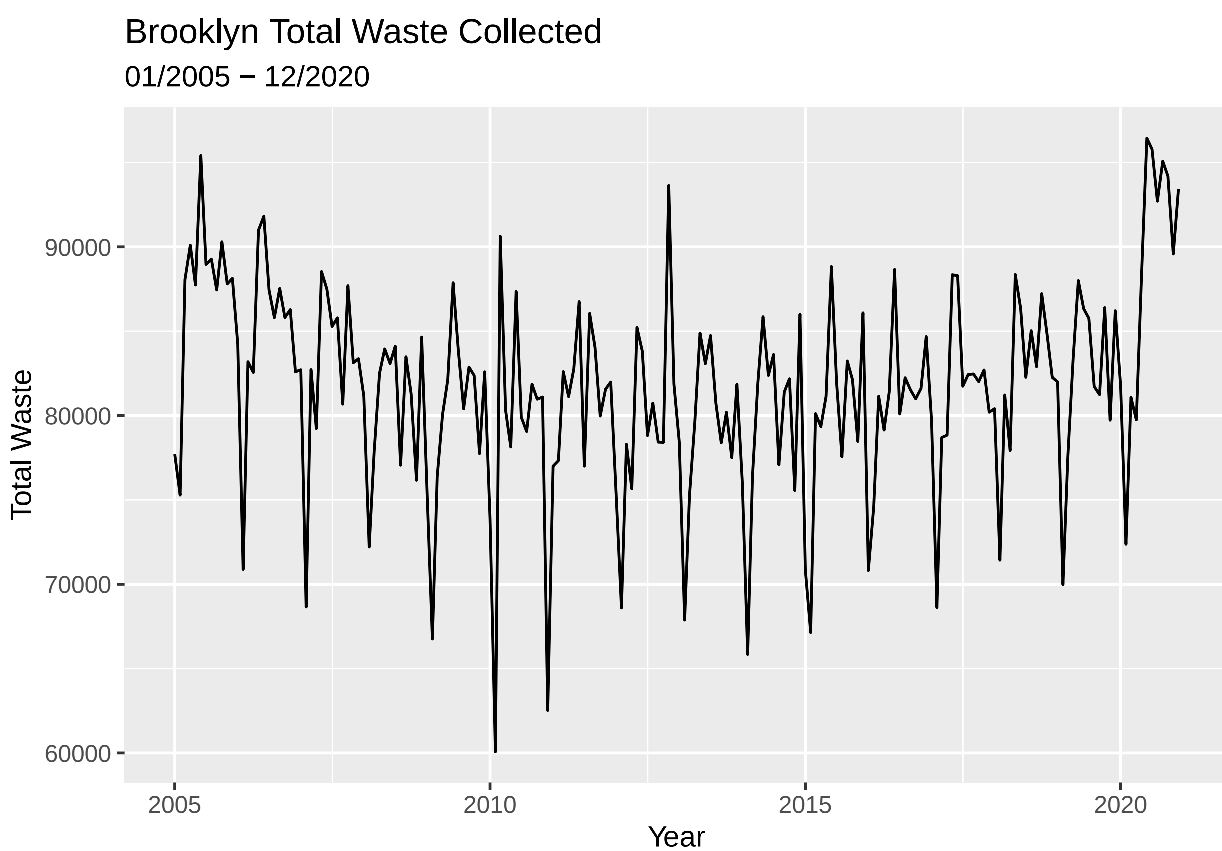
Brooklyn Total Waste

Daniel L.

5/2/2022

```r
(bk_ts) %>%
  ggplot(mapping = aes(x = (month),
                       y = total_waste)) +
  geom_line() +
  labs(x = "Year",
       y = "Total Waste",
       title = "Brooklyn Total Waste Collected",
       subtitle = "01/2005 - 12/2020")
```



The majority of the tonnage values are bounded b/w (65000,90000).

KPSS Test for 'total_waste'
H_0 : The time series is trend stationary vs H_a : The time series is not trend stationary

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$) then we reject the null hypothesis and conclude that the time series is not trend stationary.

```
#total waste values
bk_ts %>% features(total_waste, unitroot_kpss)
```

```
## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1     0.397      0.0786
```

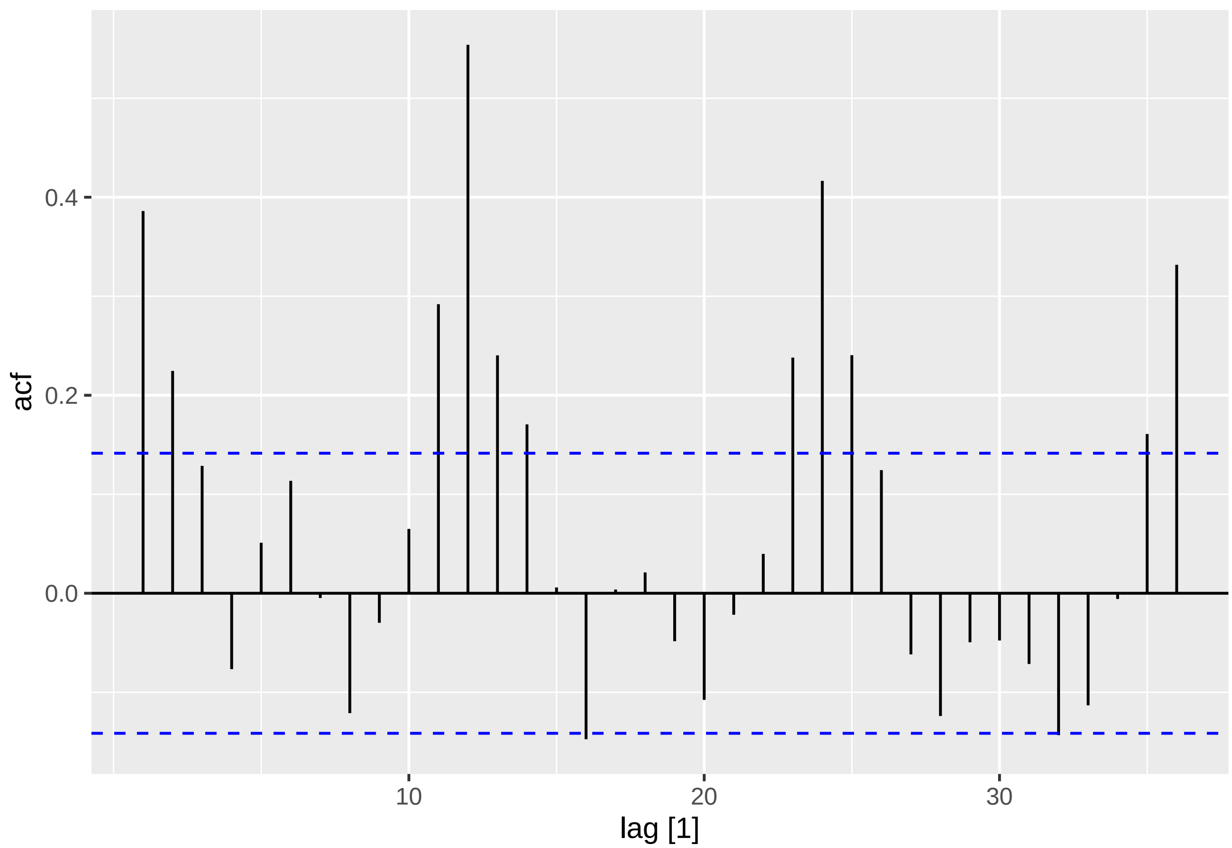
```
#differenced values
bk_ts %>% features(diff1, unitroot_kpss)
```

```
## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1    0.0278         0.1
```

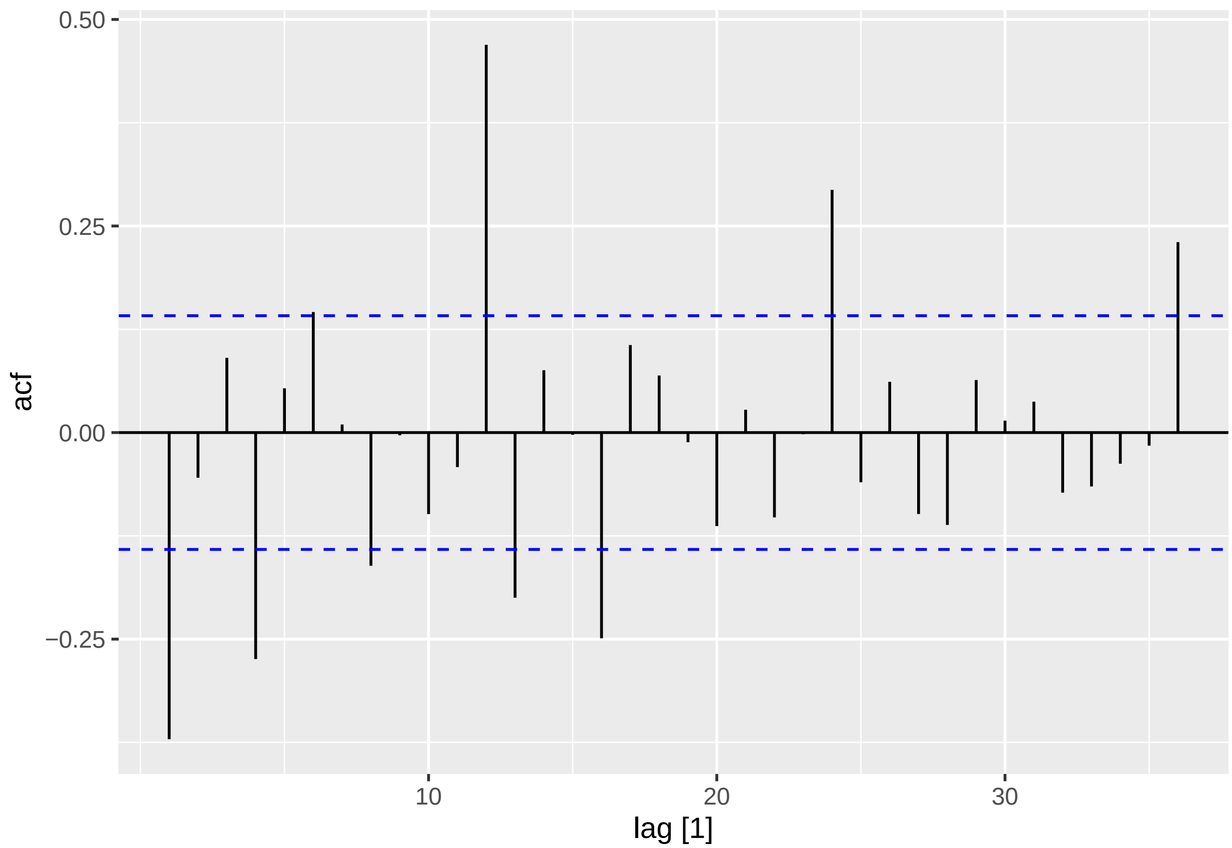
Oddly enough, the kpss test on the total waste values has us failing to reject H_0. The total waste values are trend stationary. The differenced values are also trend stationary, according the results of the kpss test.

Begin by looking at ACF and PACF of the total_waste and differenced values

```
bk_ts3 %>%
  ACF(total_waste, lag_max = 36) %>%
  autoplot()
```

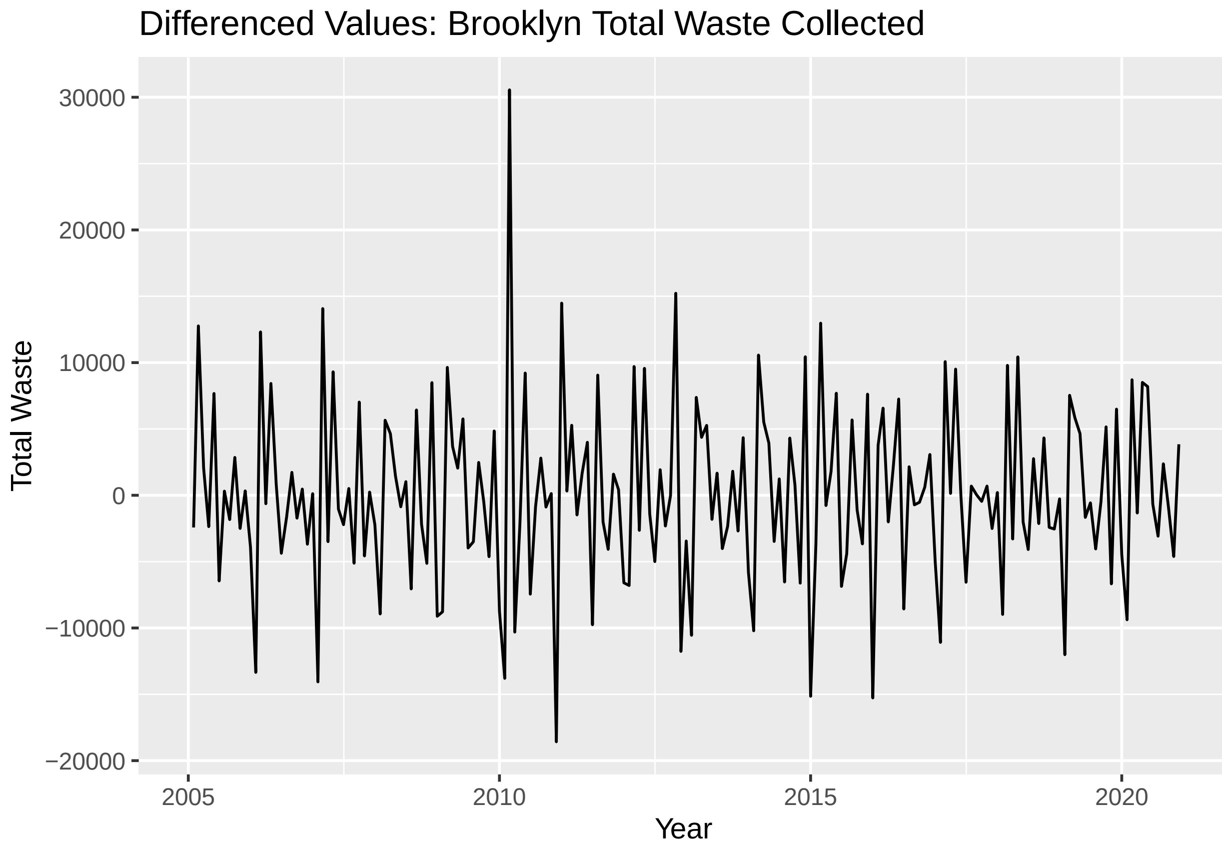



```
#acf of the differenced values
bk_ts3 %>%
  ACF(diff1, lag_max = 36) %>%
  autoplot()
```




```
bk_ts3 %>%
  ggplot(mapping = aes(x = month, y = diff1)) + geom_line() +
  labs(x = "Year", y = "Total Waste", title = "Differenced Values: Brooklyn Total Waste Collected")
```

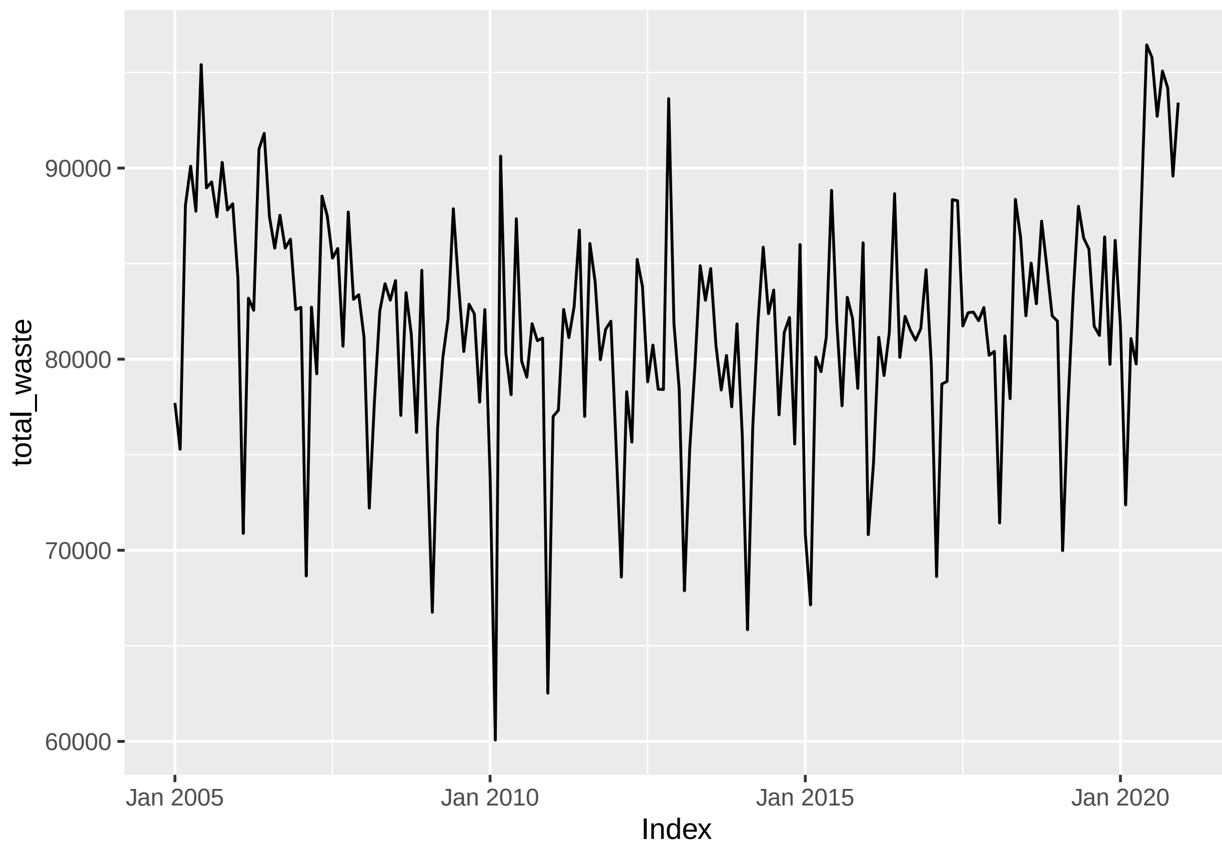
```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



Creating models with zoo() and the arima package from stats()

```
DSNY_BK_zoo_ts <- ts(DSNY_third_brooklyn[,2],
   start = as.yearmon(DSNY_third_brooklyn$month)[1],
   frequency = 12)

autoplot(as.zoo(DSNY_BK_zoo_ts))
```

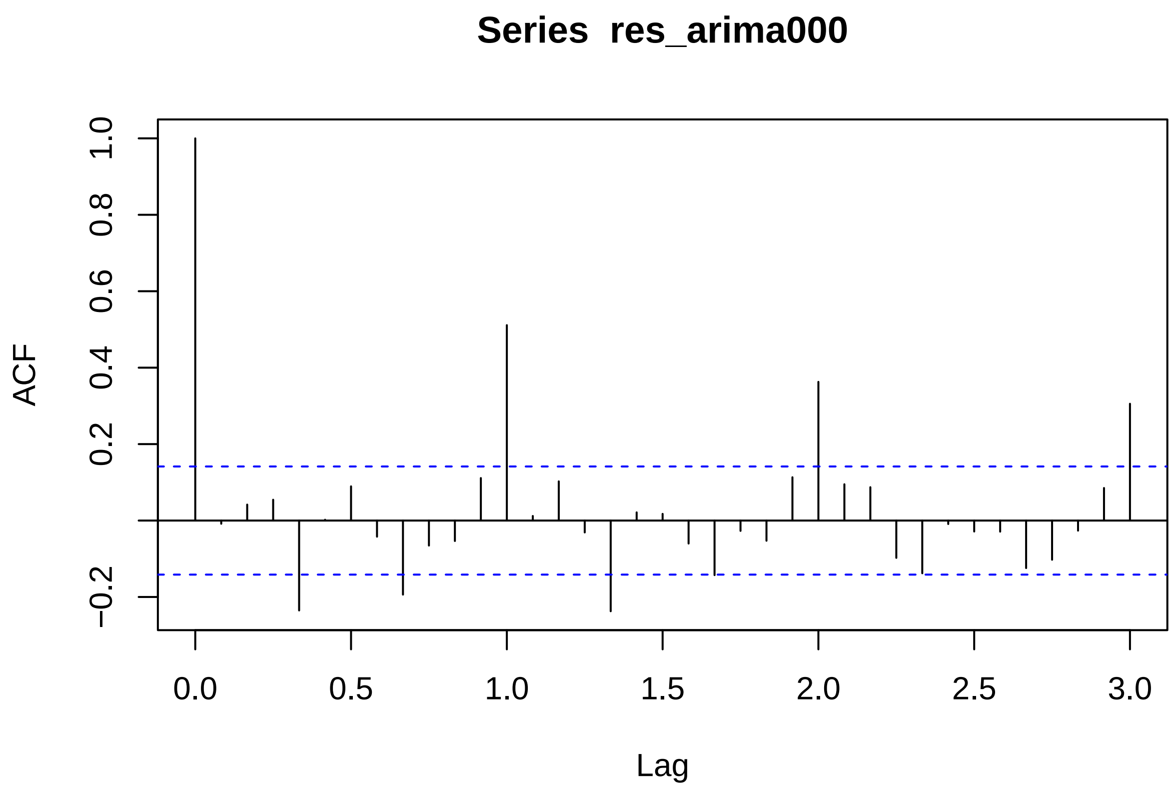


ARIMA(0,0,0) with constant

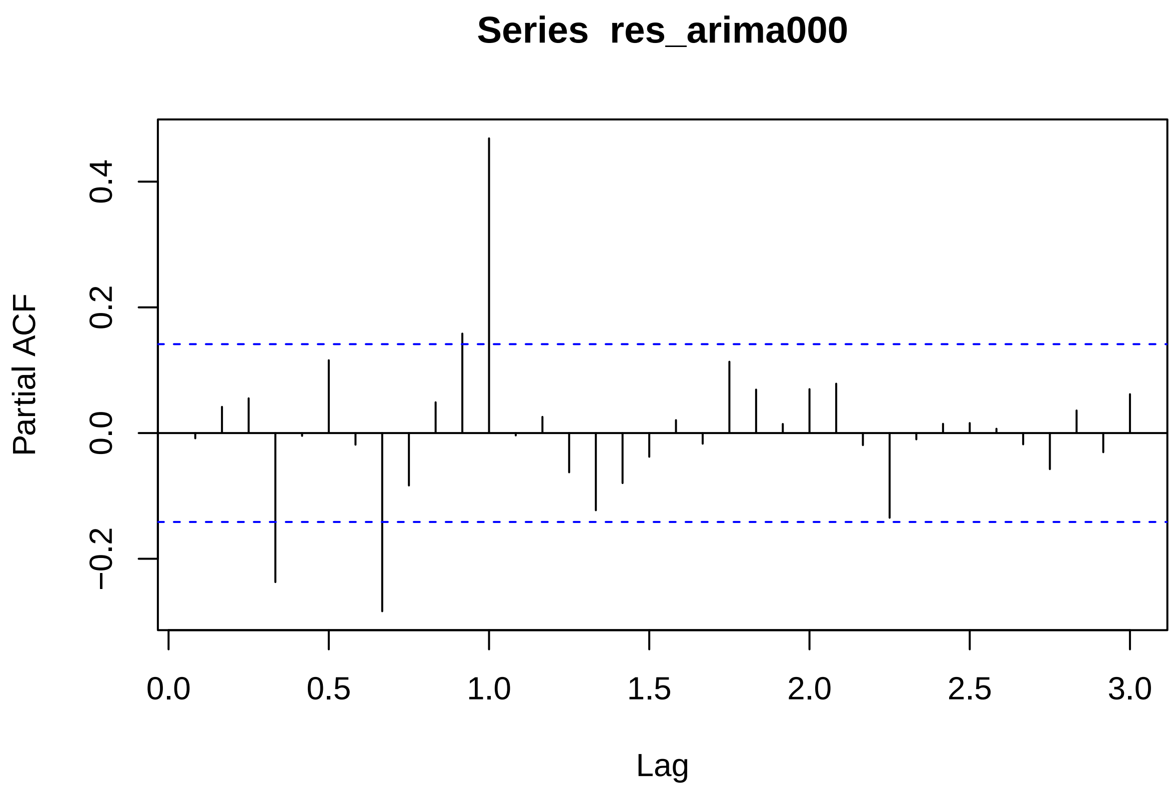
ARIMA(0,0,0)

```
bk_arima000_fit_cons <- bk_ts3 %>%
  model(arima000_constant = ARIMA(total_waste ~ 1 + pdq(0,0,0)))

zoo_arima000_fit <- arima(DSNY_BK_zoo_ts, order = 1 + c(0,0,0))
res_arima000 <- zoo_arima000_fit$residuals
acf(res_arima000, lag.max = 36)
```




```
pacf(res_arima000, lag.max = 36)
```




```
accuracy(bk_arima000_fit_cons)[4]
```

```
## # A tibble: 1 x 1
##     RMSE
##    <dbl>
## 1 6036.
```

We begin with an RMSE = 6036.26. In the ACF plot, the lags that stand out, are positive and significant are lags = (12,24,36). In the PACF plot, the first significant lag is lag = 4, which is negative. Oddly enough, lag = 12 is only seasonal lag that is positive and significant.

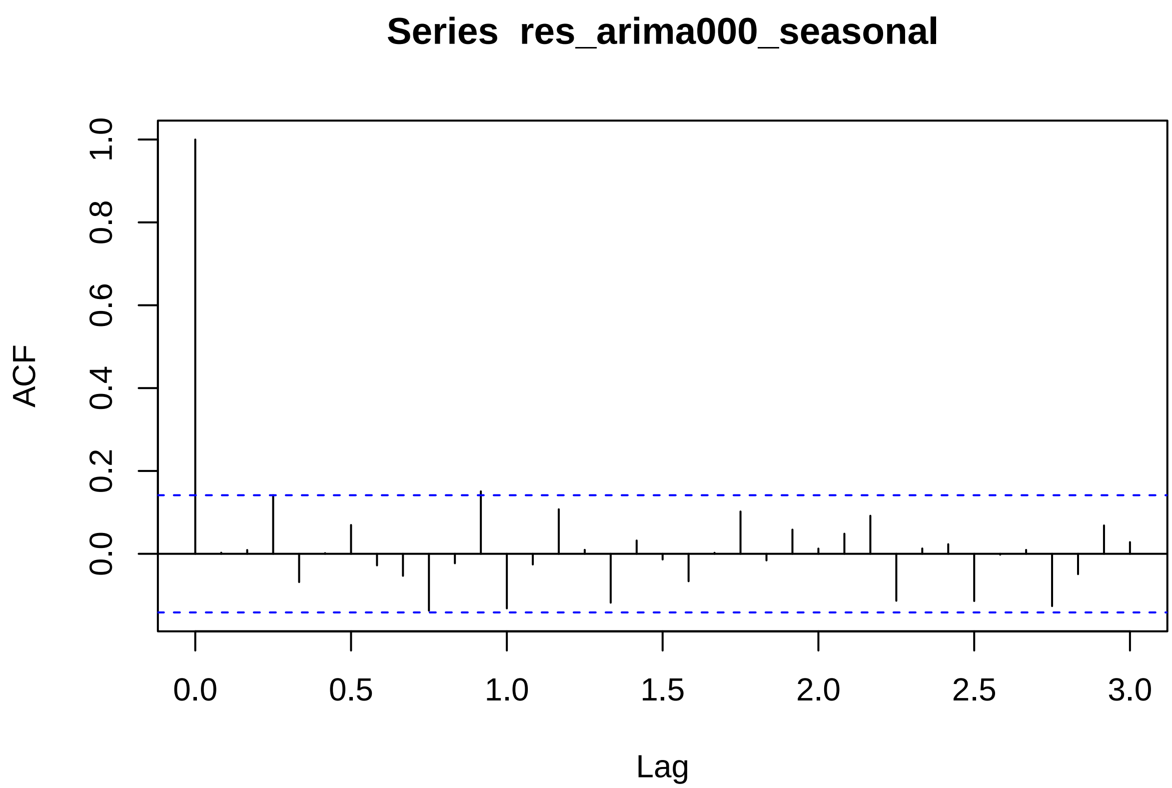
Before working with the differenced values, I will like to try different p,d,q arguments. Off the back, I would like to try adding a q = 4. But I will first try working with the seasonal parameters, at an attempt to see if the RMSE decreases

ARIMA(0,0,0)(1,0,0) with constant and seasonal parameter

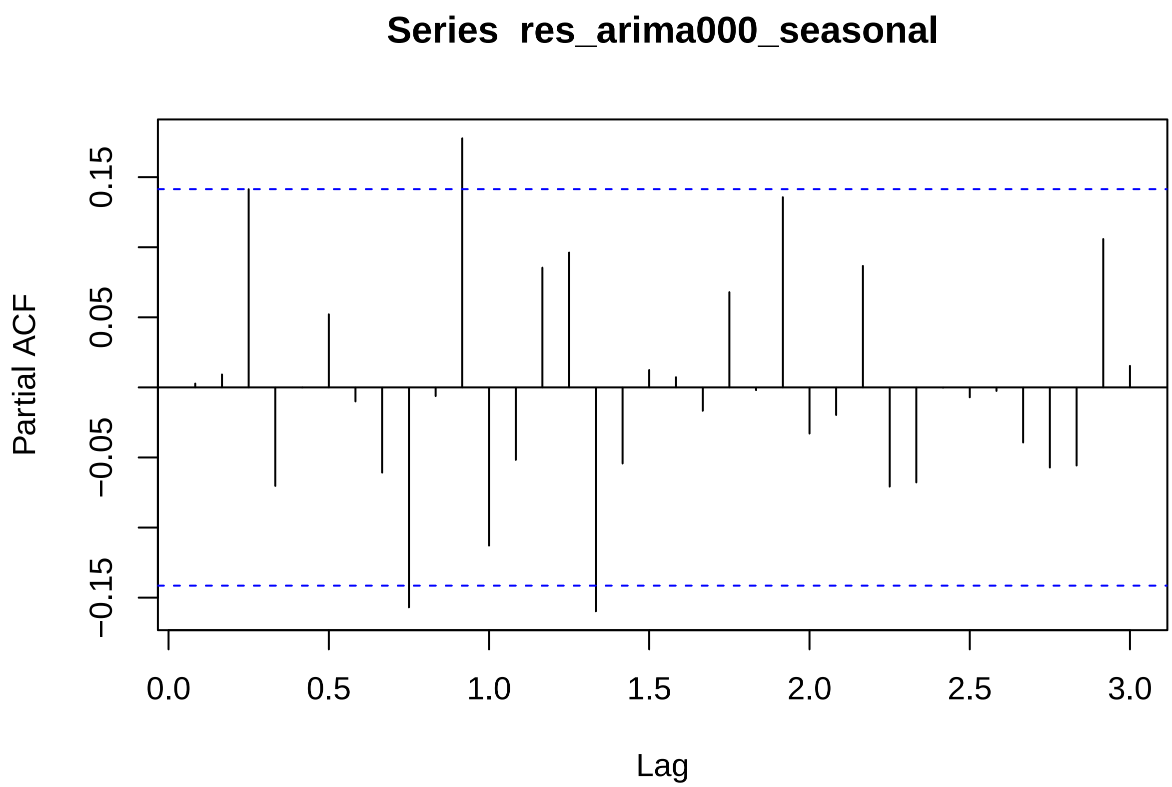
ARIMA(0,0,0)(1,0,0)

```
bk_arima000_fit_seasonal_cons <- bk_ts3 %>%
  model(arima000_constant_seasonal = ARIMA(total_waste ~ 1 +
                                     pdq(0,0,0) +
                                     PDQ(1,0,0, period = 12)))
```

```
zoo_arima000_seasonal_fit <- arima(DSNY_BK_zoo_ts,
                              order = 1 + c(0,0,0),
                              seasonal = list(order = c(1,0L,0L), period = 12))
#names(zoo_arima000_fit)
res_arima000_seasonal <- zoo_arima000_seasonal_fit$residuals
acf(res_arima000_seasonal, lag.max = 36)
```




```
pacf(res_arima000_seasonal, lag.max = 36)
```




```
accuracy(bk_arima000_fit_seasonal_cons)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 4589.
```

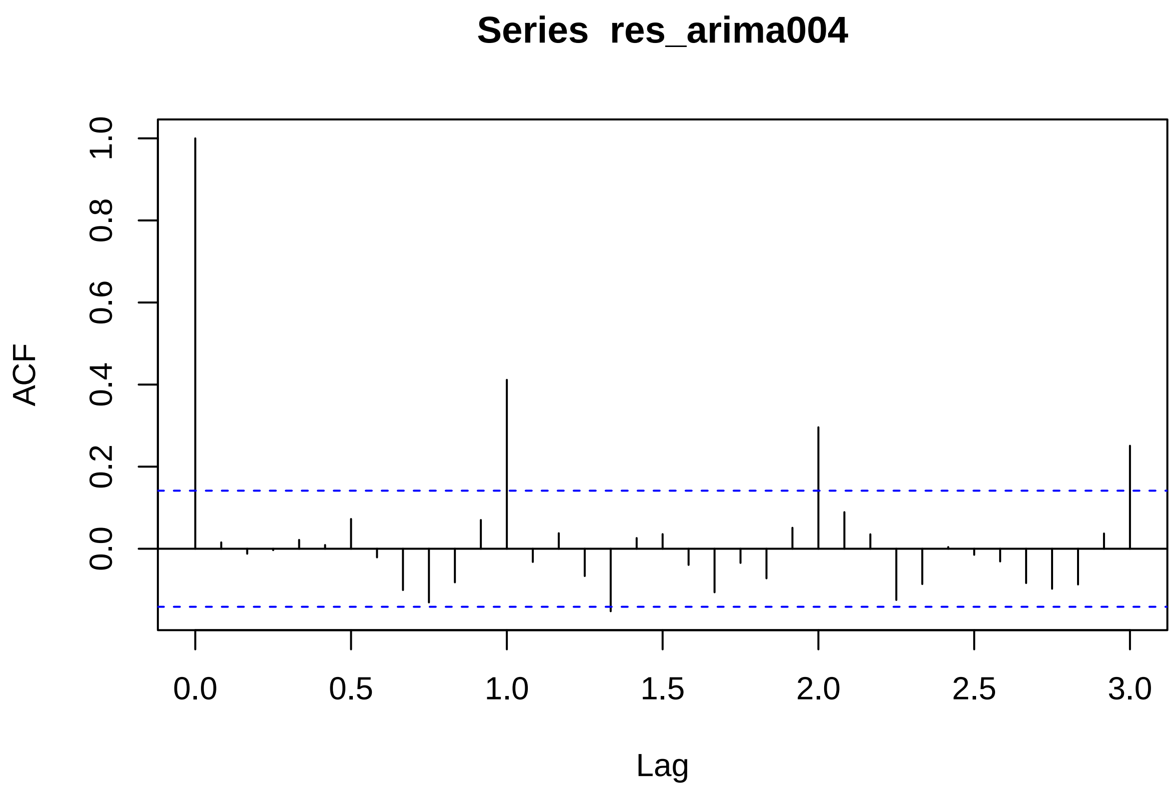
RMSE = 4589.42. The majority of the lags in the ACF plot are contained within bounds. Along with the lags of the PACF plot. Only lag = 11 is significant and positive. The values are bounded b/w (-0.15,0.15).

ARIMA(0,0,4) with constant

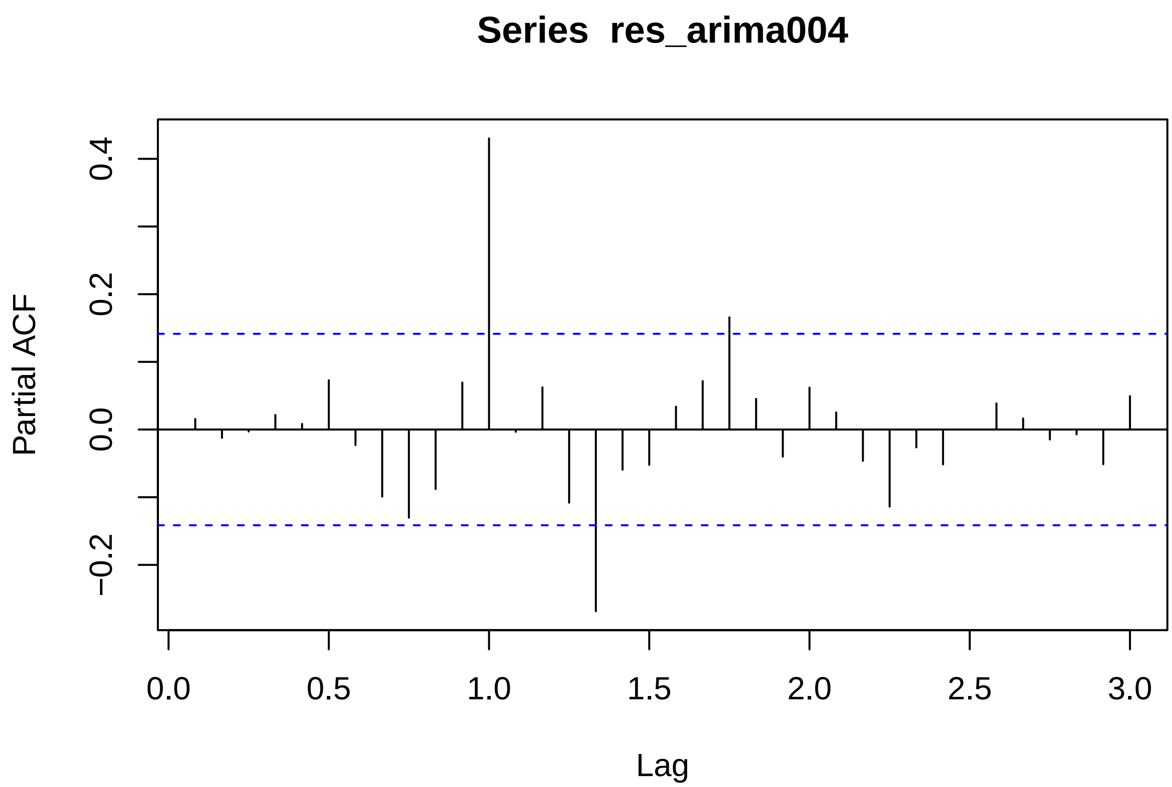
ARIMA(0,0,4)

```
bk_arima004_fit_cons <- bk_ts3 %>%
  model(arima004_constant = ARIMA(total_waste ~ 1 +
                                    pdq(0,0,4)))

zoo_arima004_fit <- arima(DSNY_BK_zoo_ts,
                          order = 1 + c(0,0,4))
#names(zoo_arima000_fit)
res_arima004 <- zoo_arima004_fit$residuals
acf(res_arima004, lag.max = 36)
```




```
pacf(res_arima004, lag.max = 36)
```




```
accuracy(bk_arima004_fit_cons)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 5325.
```

RMSE = 5325.188. We do see a decrease in the RMSE when compared to the first ARIMA(0,0,0). In both the ACF and PACF plots, the seasonal lags are significant.

Lets work with this model and add a seasonal argument.

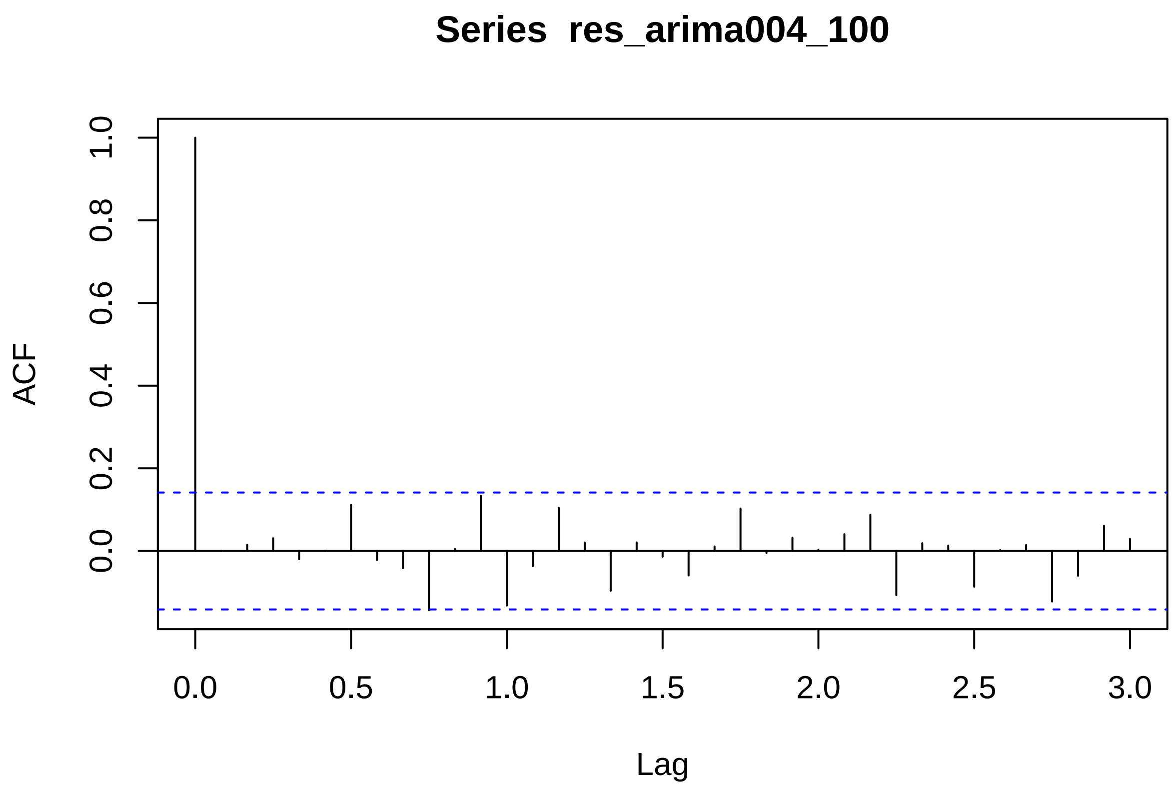
ARIMA(0,0,4)(1,0,0) with constant and seasonal

$ARIMA(0,0,4)(1,0,0)_{12}$

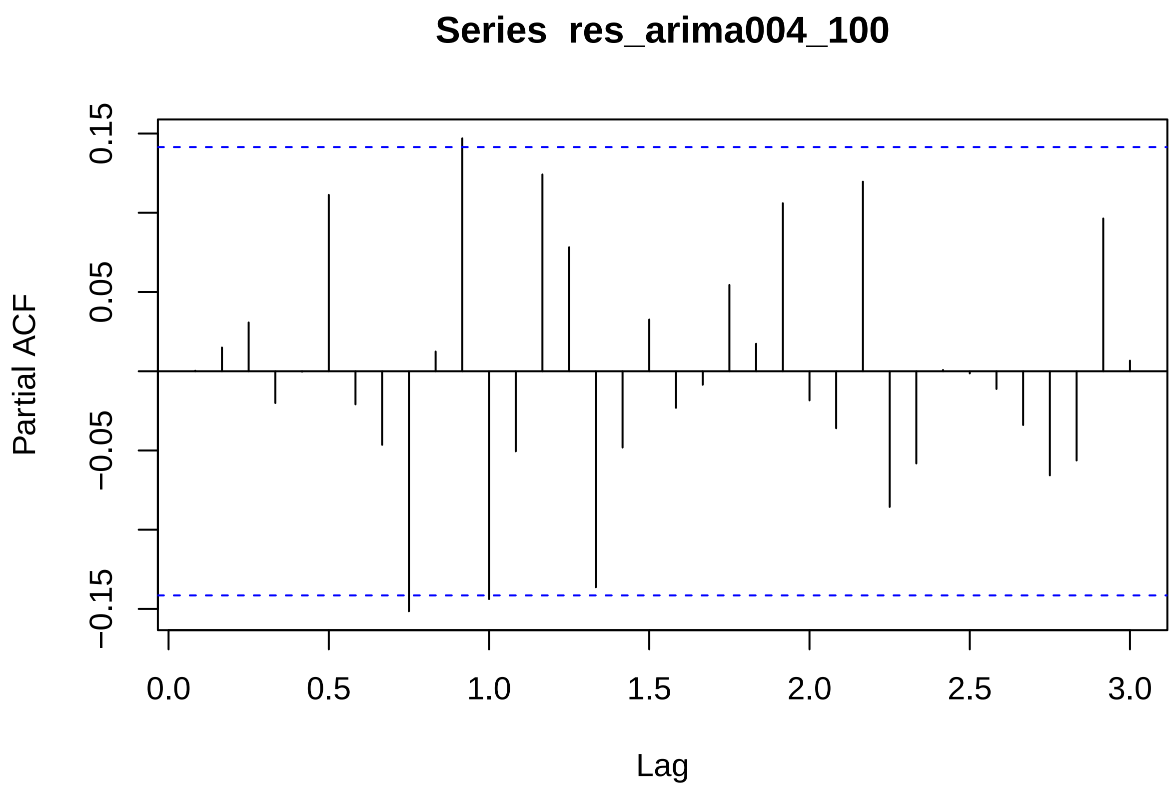
```
bk_arima004_100_seasonal_fit_cons <- bk_ts3 %>%
  model(arima004_constant = ARIMA(total_waste ~ 1 +
                                   pdq(0,0,4) +
                                   PDQ(1,0,0,
                                       period = 12)))

zoo_arima004_100_seasonalfit <- arima(DSNY_BK_zoo_ts,
                          order = 1 + c(0,0,4),
```

```
                                seasonal = list(order = c(1, 0L, 0L), period = 12))
#names(zoo_arima000_fit)
res_arima004_100 <- zoo_arima004_100_seasonalfit$residuals
acf(res_arima004_100, lag.max = 36)
```




```
pacf(res_arima004_100, lag.max = 36)
```




```
accuracy(bk_arima004_100_seasonal_fit_cons)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 4445.
```

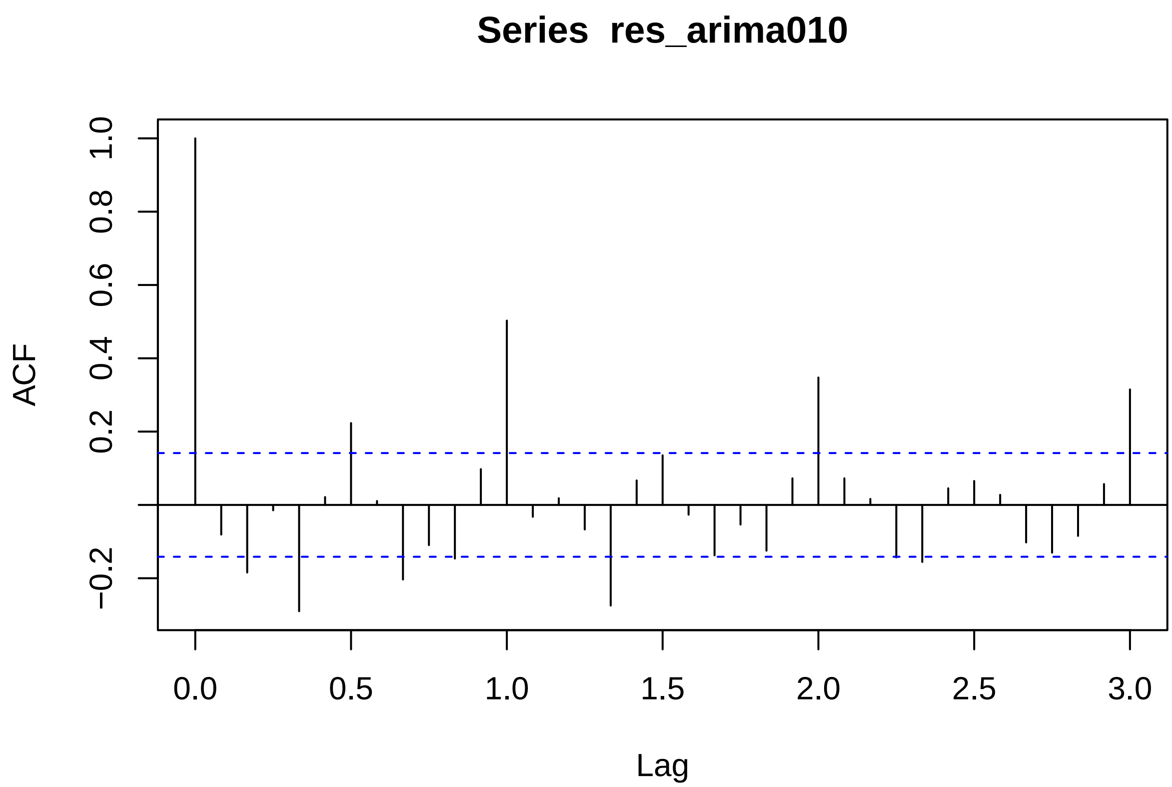
RMSE = 4444.908. Most of the lagged values are not-significant. They look good and appear to be bounded. Since we can also work with the differenced values, we will create some models with them

$ARIMA(0,1,0)$ with constant

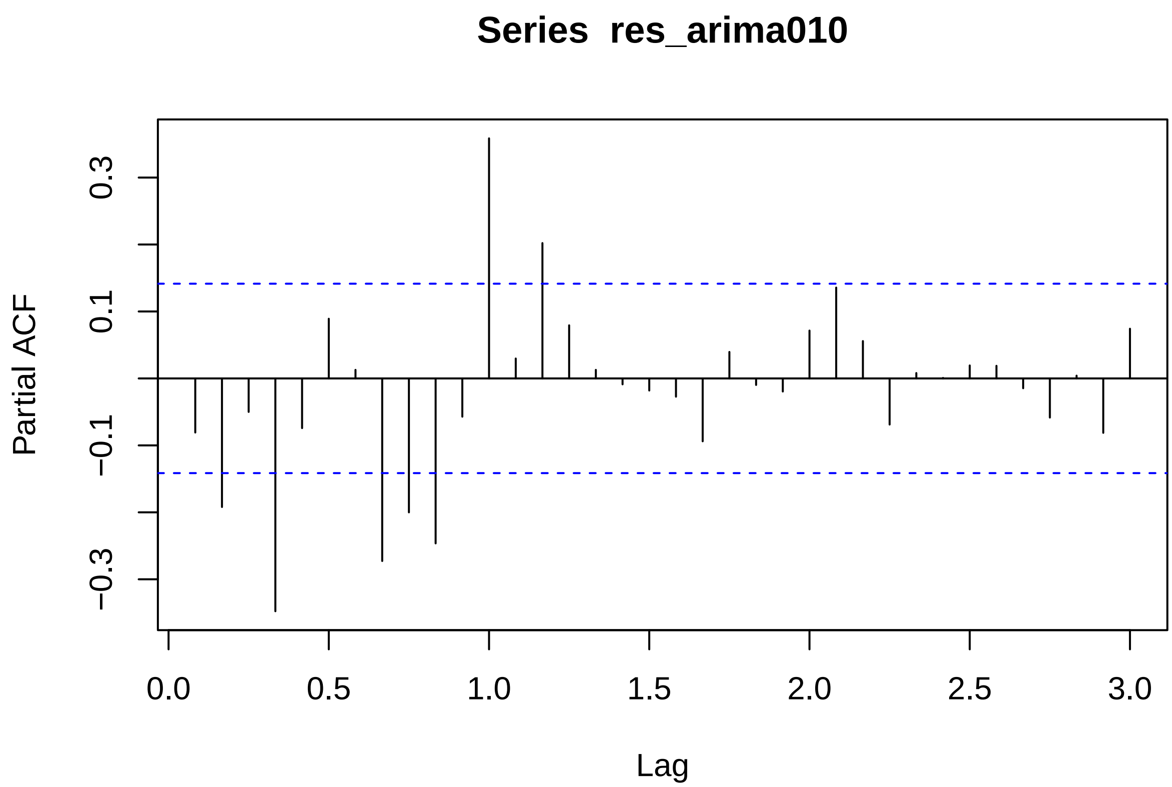
$ARIMA(0,1,0)$

```
bk_arima010_fit_cons <- bk_ts3 %>%
  model(arima010_constant = ARIMA(total_waste ~ 1 + pdq(0,1,0)))

zoo_arima010_fit <- arima(DSNY_BK_zoo_ts, order = 1 + c(0,1,0))
res_arima010 <- zoo_arima010_fit$residuals
acf(res_arima010, lag.max = 36)
```




```r
pacf(res_arima010, lag.max = 36)
```




```
accuracy(bk_arima010_fit_cons)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 6629.
```

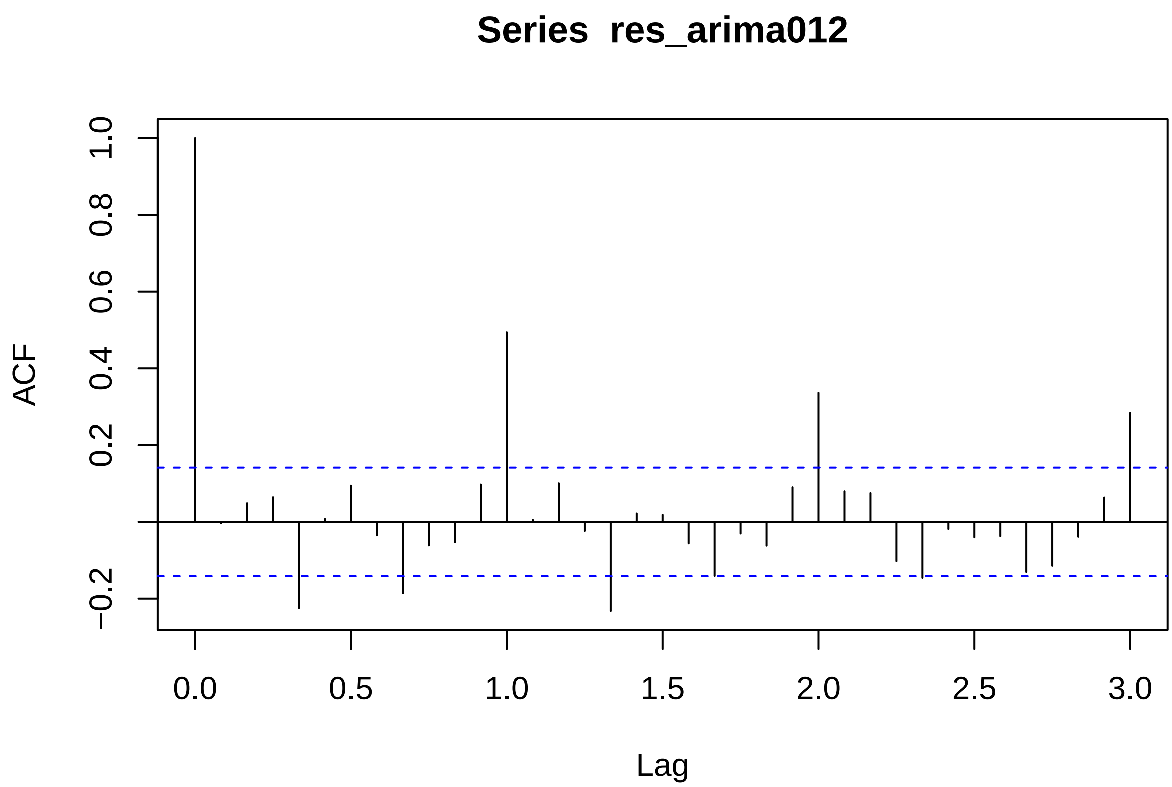
RMSE = 6628.688. Try q = 2 or q = 4

$ARIMA(0,1,2)$ with constant

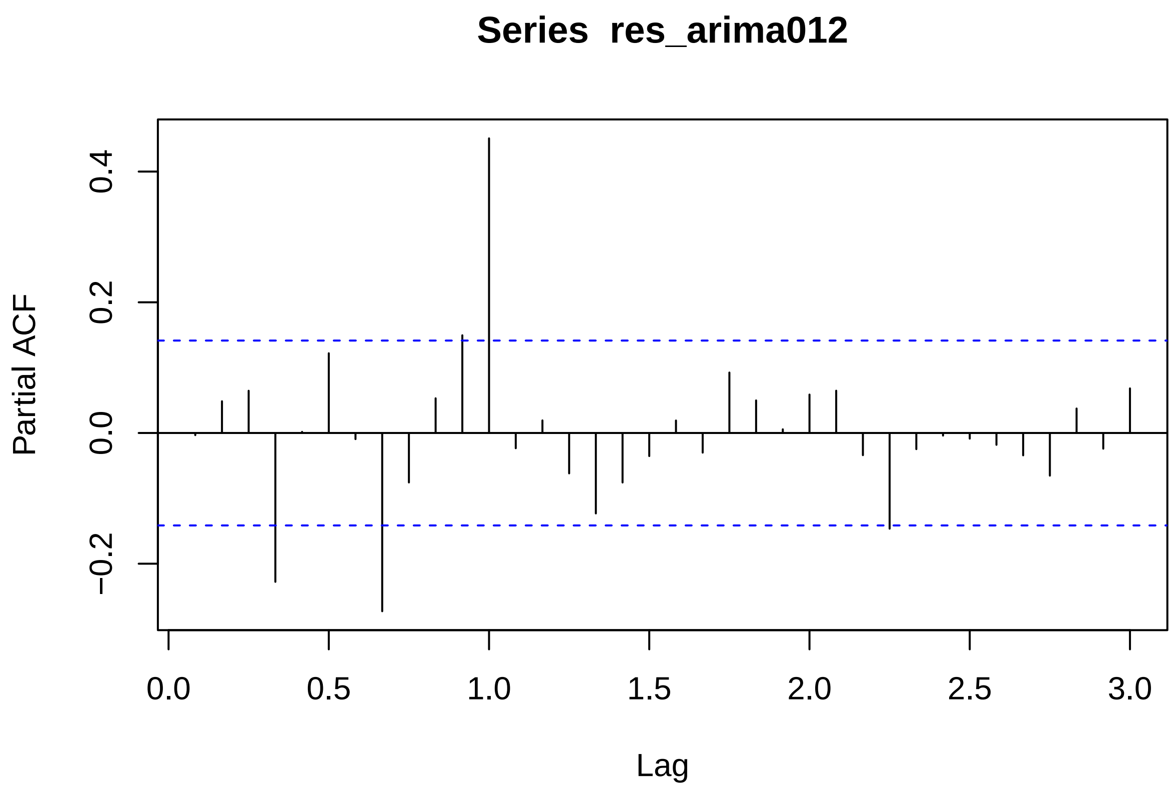
$ARIMA(0,1,2)$

```
bk_arima012_fit_cons<- bk_ts3 %>%
  model(arima012_constant = ARIMA(total_waste ~ 1 + pdq(0,1,2)))

zoo_arima012_fit <- arima(DSNY_BK_zoo_ts, order = 1 + c(0,1,2))
res_arima012 <- zoo_arima012_fit$residuals
acf(res_arima012, lag.max = 36)
```




```
pacf(res_arima012, lag.max = 36)
```




```
accuracy(bk_arima012_fit_cons)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 5529.
```

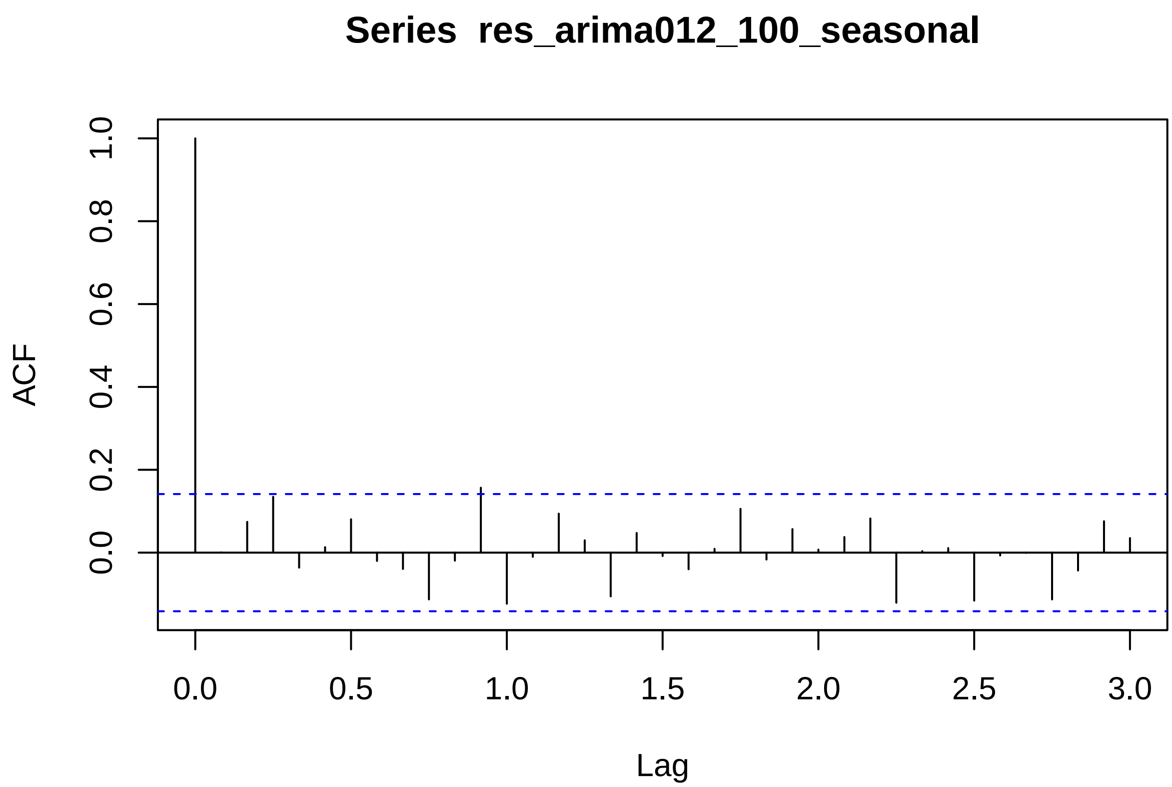
RMSE = 5528.971. Lag 4 in the PACF is still significant, along with the seasonal lag.

$ARIMA(0,1,2)(1,0,0)$ with constant

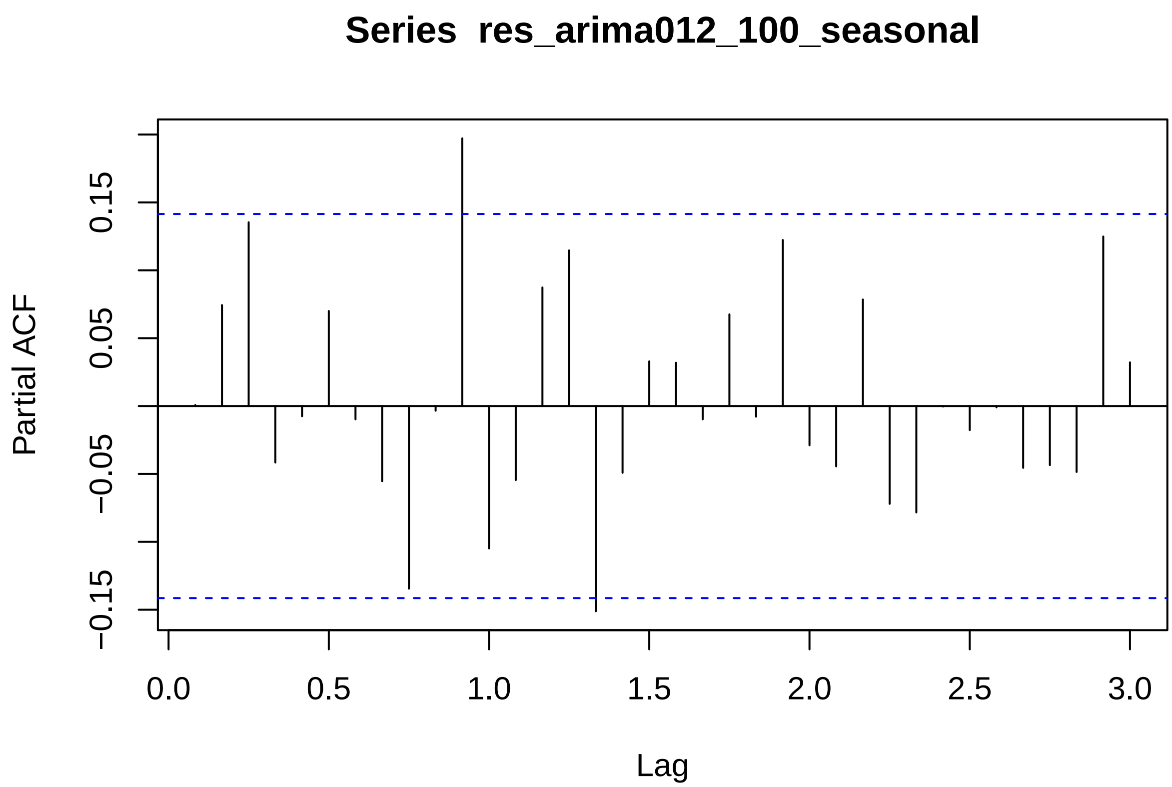
$ARIMA(0,1,2)(1,0,0)_{12}$

```
bk_arima012_100_fit_seasonal_cons <- bk_ts3 %>%
  model(arima012_100_constant_seasonal = ARIMA(total_waste ~ 1 +
                                     pdq(0,1,2) +
                                     PDQ(1,0,0, period = 12)))

zoo_arima012_100_seasonal_fit <- arima(DSNY_BK_zoo_ts,
                          order = 1 + c(0,1,2),
                          seasonal = list(order = c(1,0L,0L), period = 12))
#names(zoo_arima000_fit)
res_arima012_100_seasonal <- zoo_arima012_100_seasonal_fit$residuals
acf(res_arima012_100_seasonal, lag.max = 36)
```




```
pacf(res_arima012_100_seasonal, lag.max = 36)
```




```
accuracy(bk_arima012_100_fit_seasonal_cons)[4]
```

```
## # A tibble: 1 x 1
##     RMSE
##    <dbl>
## 1 4427.
```

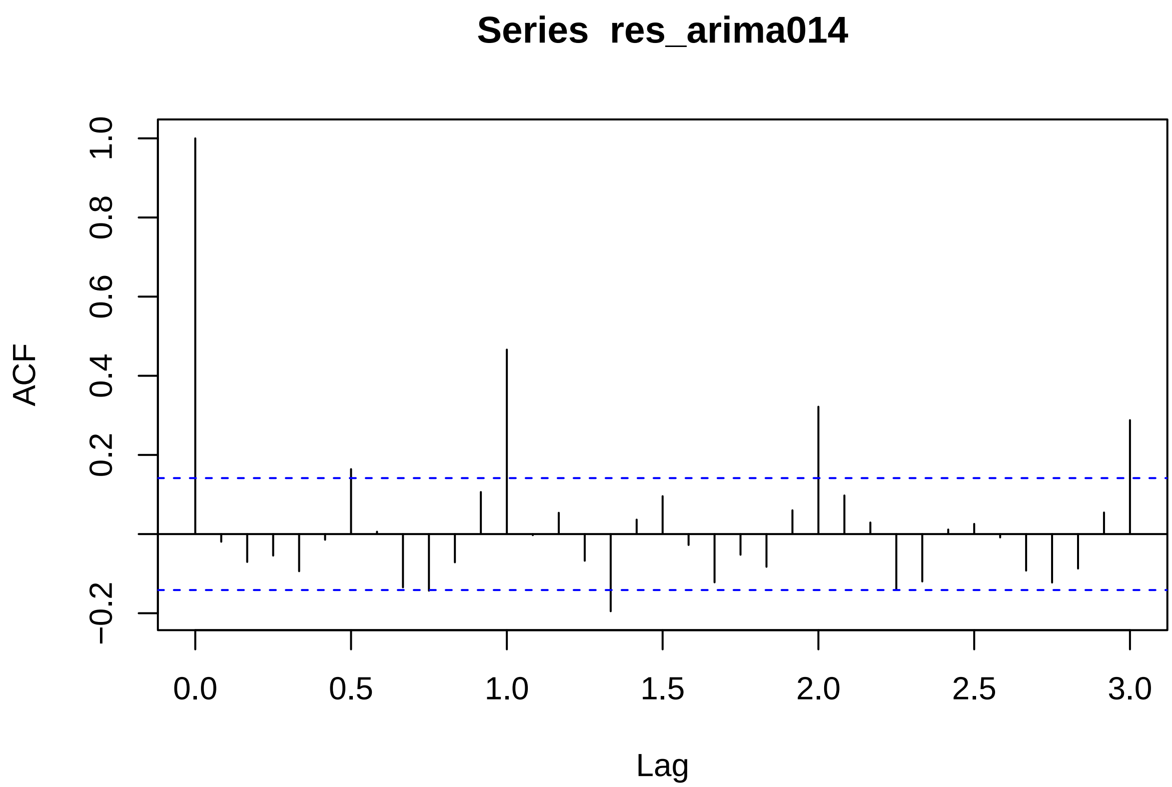
RMSE $= 4427.169$

$ARIMA(0,1,4)$ with constant

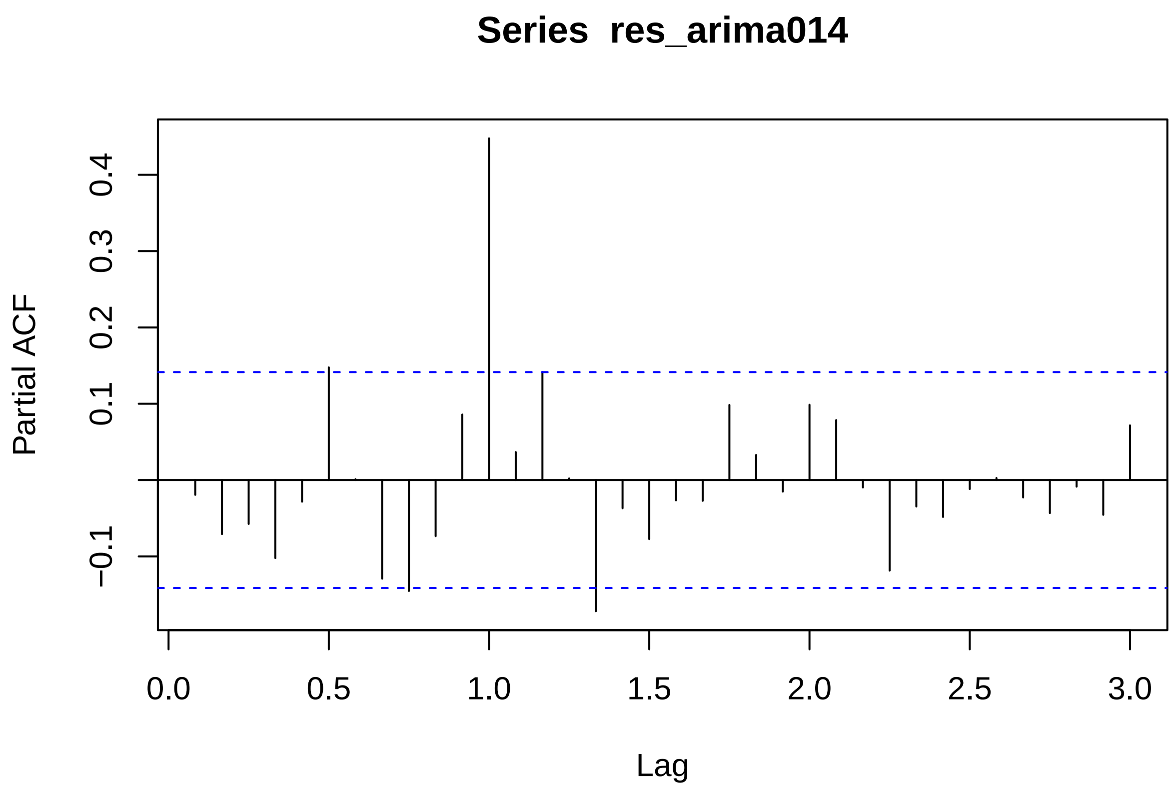
$ARIMA(0,1,4)$

```
bk_arima014_fit_cons<- bk_ts3 %>%
  model(arima014_constant = ARIMA(total_waste ~ 1 + pdq(0,1,4)))

zoo_arima014_fit <- arima(DSNY_BK_zoo_ts, order = 1 + c(0,1,4))
res_arima014 <- zoo_arima014_fit$residuals
acf(res_arima014, lag.max = 36)
```




```
pacf(res_arima014, lag.max = 36)
```




```
accuracy(bk_arima014_fit_cons)[4]
```

```
## # A tibble: 1 x 1
##     RMSE
##    <dbl>
## 1 5410.
```

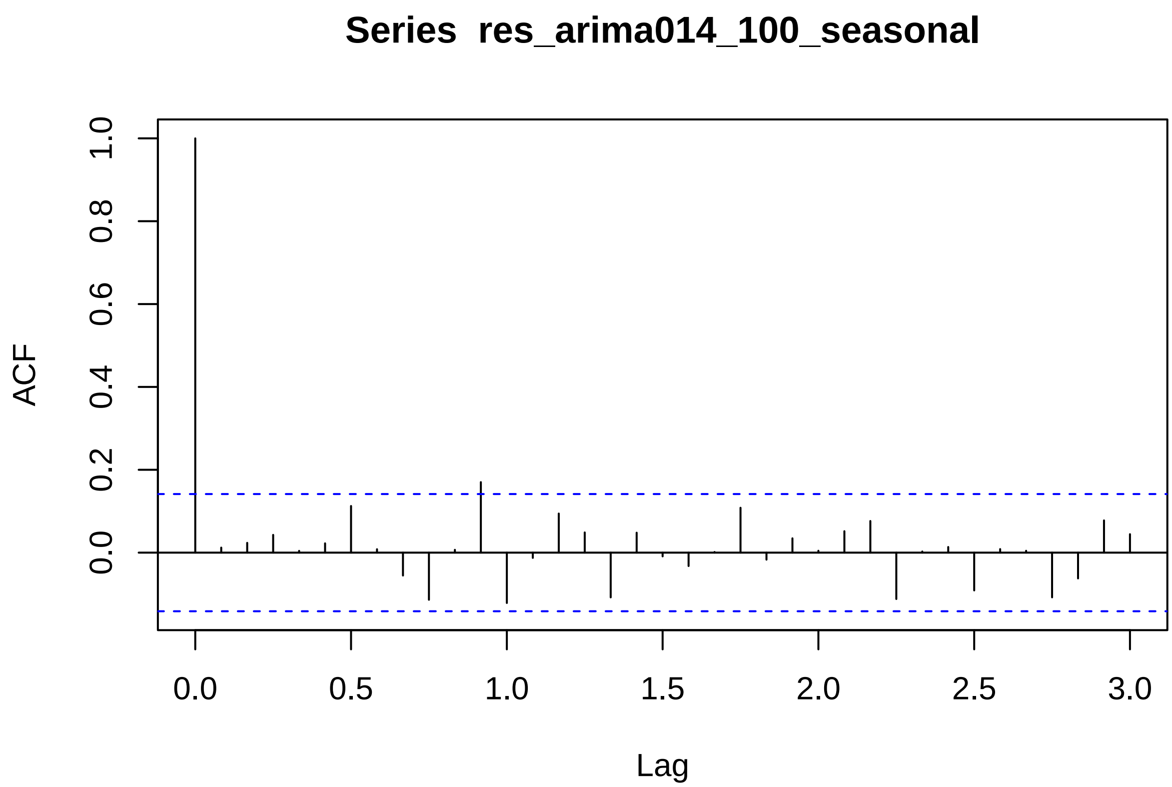
RMSE $= 5410.265$

$ARIMA(0,1,4)(1,0,0)$ with constant

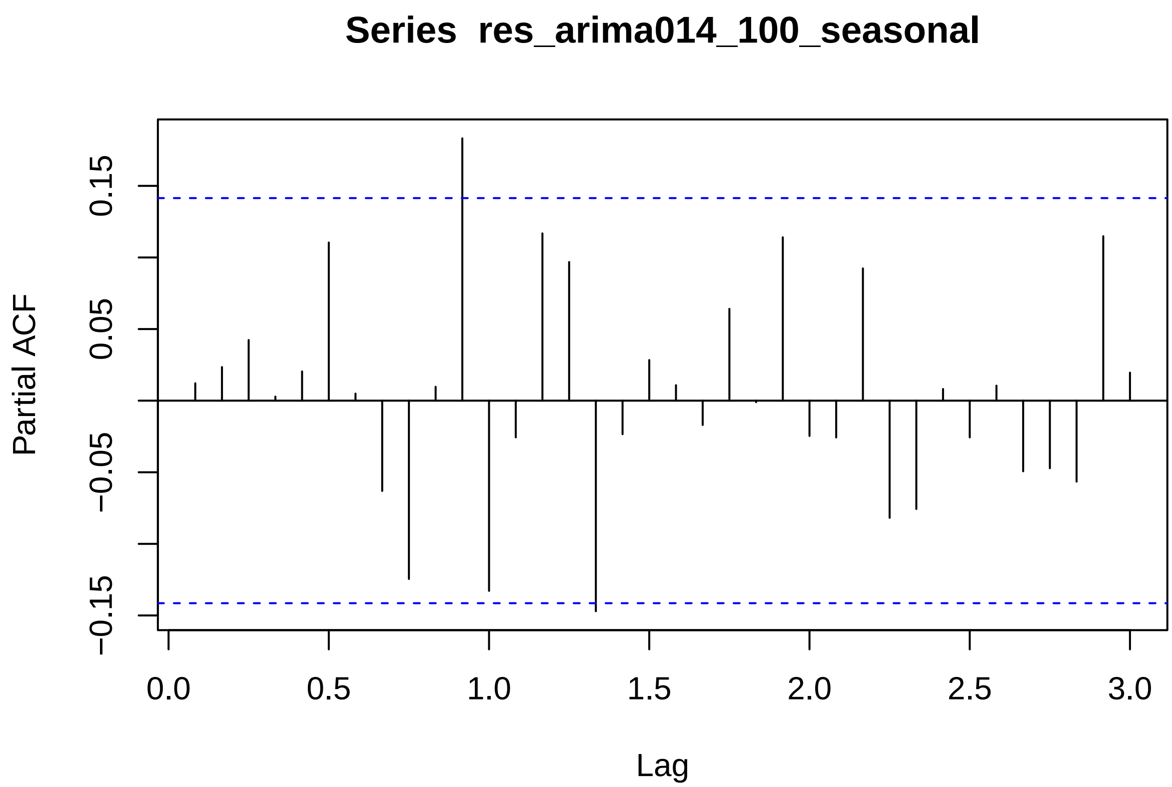
$ARIMA(0,1,4)(1,0,0)_{12}$

```
bk_arima014_100_fit_seasonal_cons <- bk_ts3 %>%
  model(arima014_100_constant_seasonal = ARIMA(total_waste ~ 1 +
                                   pdq(0,1,4) +
                                   PDQ(1,0,0, period = 12)))

zoo_arima014_100_seasonal_fit <- arima(DSNY_BK_zoo_ts,
                          order = 1 + c(0,1,4),
                          seasonal = list(order = c(1,0L,0L), period = 12))
#names(zoo_arima000_fit)
res_arima014_100_seasonal <- zoo_arima014_100_seasonal_fit$residuals
acf(res_arima014_100_seasonal, lag.max = 36)
```




```
pacf(res_arima014_100_seasonal, lag.max = 36)
```




```
accuracy(bk_arima014_100_fit_seasonal_cons)[4]
```

```
## # A tibble: 1 x 1
##     RMSE
##    <dbl>
## 1 4393.
```

RMSE = 4392.534

Auto-arima

For our final models, we will look and compare the results of an auto-arima model from the feasts package.

```
bk_auto_arima_fit_cons <- bk_ts3 %>%
  model(stepwise = ARIMA(total_waste),
        search = ARIMA(total_waste,
                       stepwise = FALSE,
                       approximation = FALSE))
accuracy(bk_auto_arima_fit_cons)[1:4]
```

```
## # A tibble: 2 x 4
##   .model   .type       ME  RMSE
##   <chr>    <chr>    <dbl> <dbl>
## 1 stepwise Training  15.6 5257.
## 2 search   Training  14.2 5220.
```

```r
bk_auto_arima_fit_cons %>% select(.model = stepwise) %>% report()
```

```
## Series: total_waste
## Model: ARIMA(2,0,3) w/ mean
##
## Coefficients:
##          ar1      ar2      ma1     ma2     ma3    constant
##       0.4450  -0.5832  -0.0805  0.7556  0.2529  93085.8012
## s.e.  0.3268   0.1768   0.3578  0.0771  0.1785    729.6156
##
## sigma^2 estimated as 28523169:  log likelihood=-1917.85
## AIC=3849.7   AICc=3850.31   BIC=3872.5
```

```r
print("--------------")
```

```
## [1] "--------------"
```

```r
bk_auto_arima_fit_cons %>% select(.model = search) %>% report()
```

```
## Series: total_waste
## Model: ARIMA(3,0,3) w/ mean
##
## Coefficients:
##          ar1      ar2      ar3     ma1     ma2     ma3     constant
##       0.0420  -0.2942  -0.3053  0.3209  0.5731  0.6015  127365.4131
## s.e.  0.1799   0.1428   0.1524  0.1594  0.0962  0.1200     938.3475
##
## sigma^2 estimated as 28276150:  log likelihood=-1916.54
## AIC=3849.08   AICc=3849.87   BIC=3875.14
```

Summary of models

$ARIMA(0,0,4)(1,0,0)_{12}$ has $RMSE = 4444.908$
$ARIMA(0,1,2)(1,0,0)_{12}$ has $RMSE = 4427.169$
$ARIMA(0,1,4)(1,0,0)_{12}$ has $RMSE = 4392.534$
$ARIMA(3,0,3)$ has $RMSE = 5219.698$