HW3

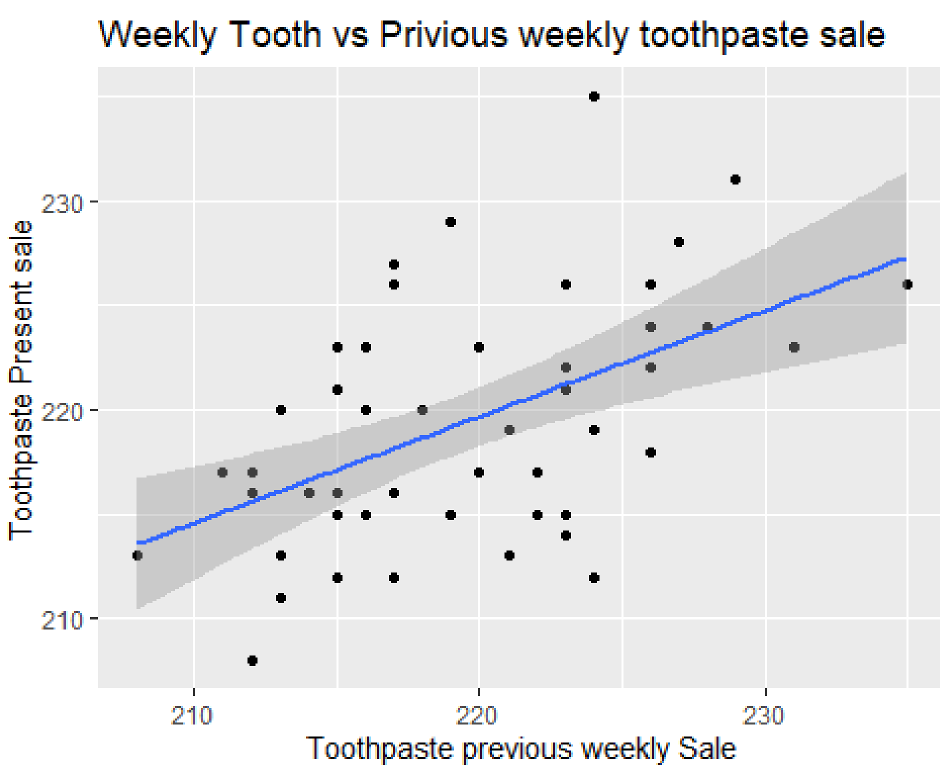
Oluwafemi Shobowale

4/18/2022

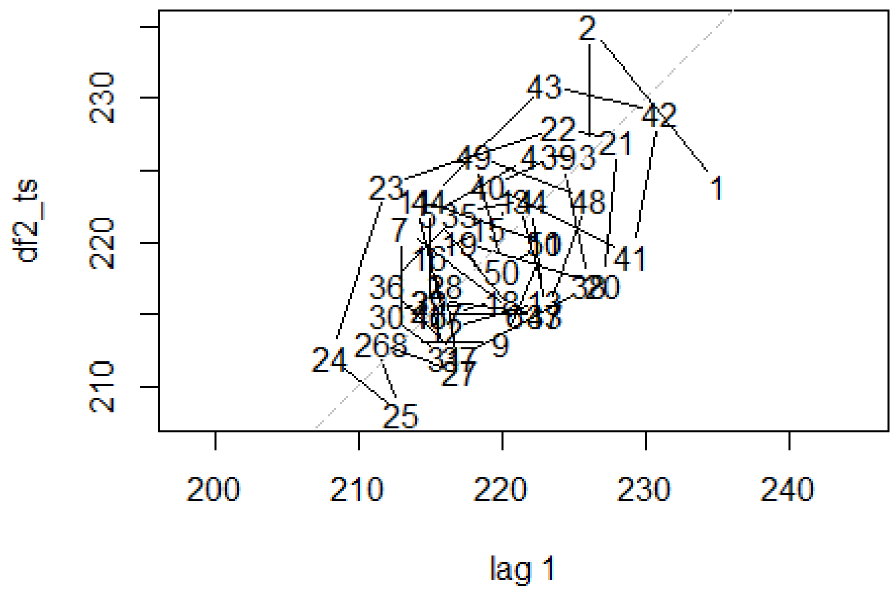
1,

1a, There seem to be no serial correlation between the lag weekly toothpaste sale and present tooth paste sale.

```
ggplot(df2,aes(y= ToothPaste ,x=ToothPaste_lag
,group=1))+geom_point()+geom_smooth(method = lm)+labs(title = "Weekly Tooth
vs Privious weekly toothpaste sale")+ylab("Toothpaste Present
sale")+xlab("Toothpaste previous weekly Sale")
```

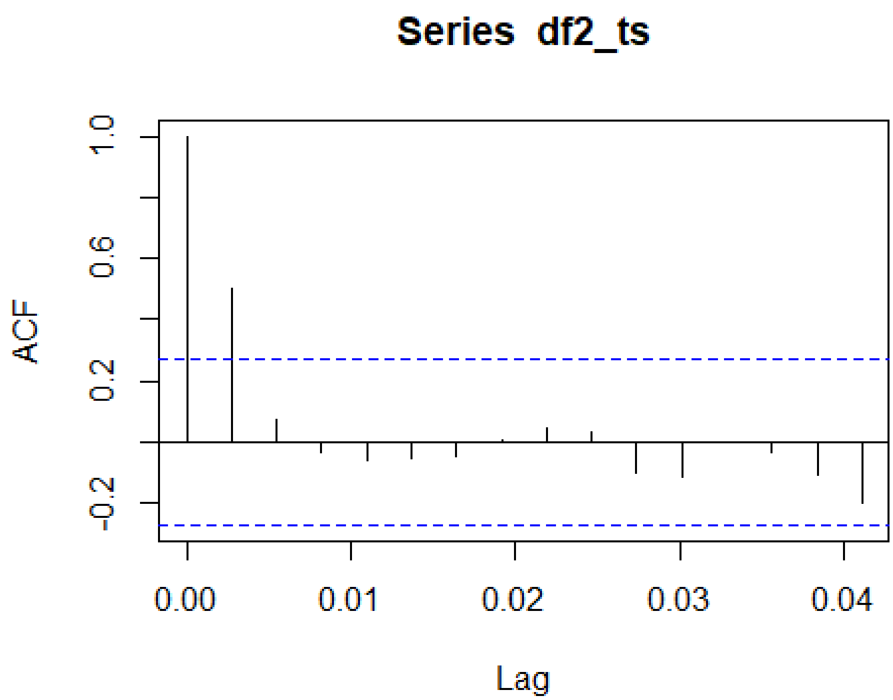



```
lag.plot(df2_ts, lags = 1)
```



1b, There is evidence of serial correlation seen in the ACF graph bellow since there is a signicicant auto correlation at lag=1.

```
acf(df2_ts, lag.max = 15)
```



1c, We fail to reject the null hypotesis series series since p-value > α. Hence we suspect some sort of correlation in the series.

```
#Ljung-Box test
Box.test(df2_ts,lag=10, type = "Ljung-Box")

##
##  Box-Ljung test
##
## data:  df2_ts
## X-squared = 15.819, df = 10, p-value = 0.1049
```

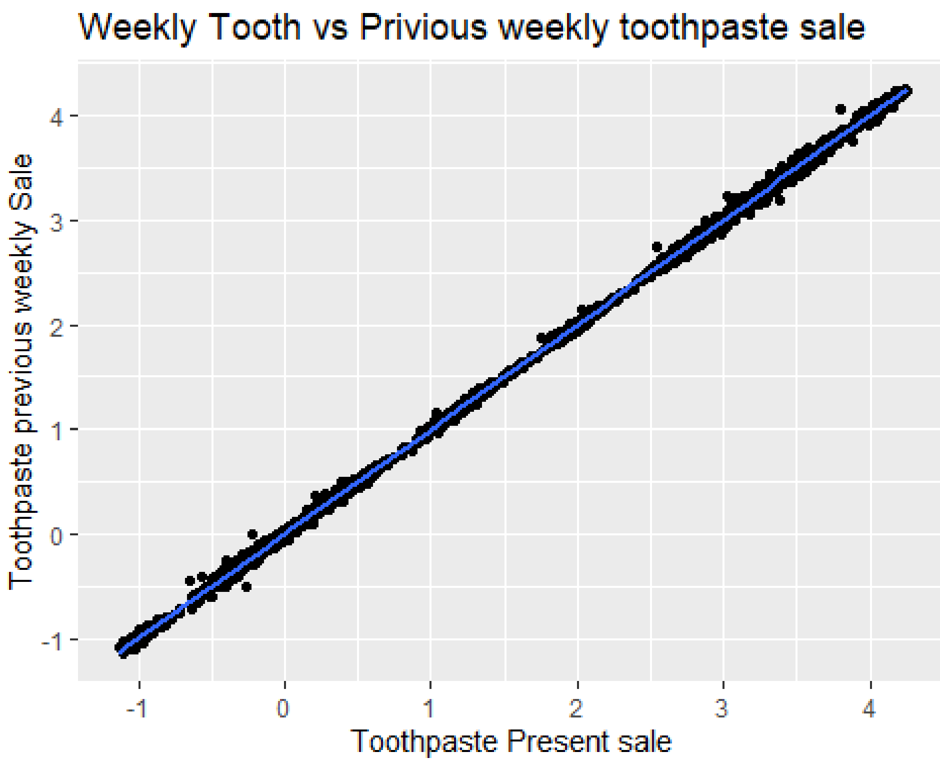
1d, Weak stationarity implies common mean and variance in a series, time invarient and auto-correlation does not depend on time. The practical implication is that mean and variance are the same in the past and future. One way to detect staionarity is to plot the time series and notice if a trend is visible over time or if variance change.

2

2a, There seem to be a strong positive linear correlation between the log price and lag_log price. Auto correlation is suspected in this series.

```
ggplot(df3,aes(x= log_price ,y= lag_price
,group=1))+geom_point()+geom_smooth(method = lm)+labs(title = "Weekly Tooth
```

```
vs Privious weekly toothpaste sale")+xlab("Toothpaste Present
sale")+ylab("Toothpaste previous weekly Sale")
```



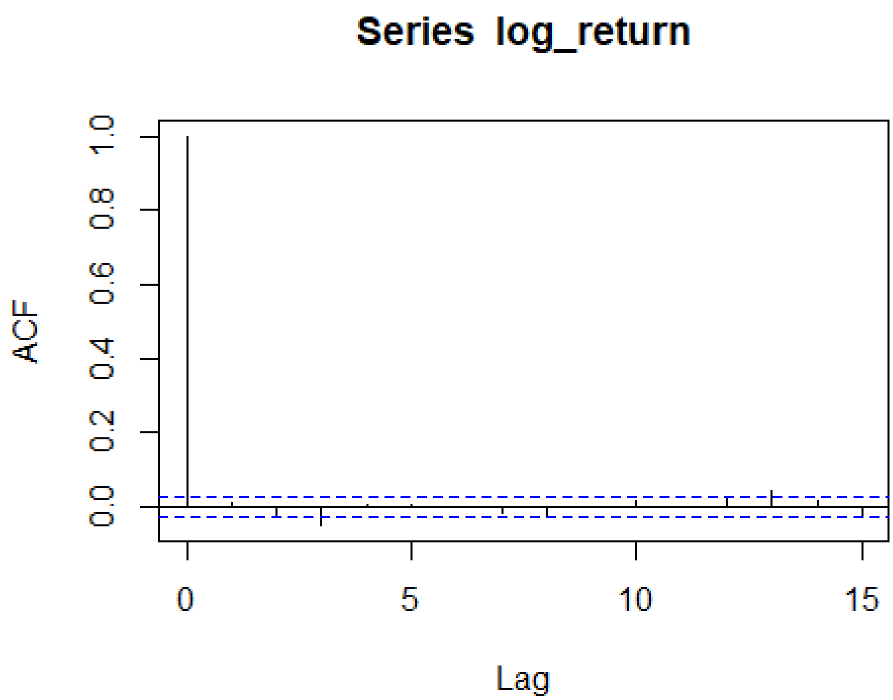
2b, The ACF shows a high correlation at all 15 lags. There is a strong evidence of serial correlation in this series.

```
acf(df3$log_price, lag.max = 15)
```

2c, The first 10 lags are are are all constant at one. ACF shows the series is a non stationary series because the ACF slowly falls off in this case the ACF is constant along all lags.

2d, Computing the log return ln(price)-ln(lag_price). Log retun has a zero auto correlation along all lags. There seems to be no dependency in the the log returns,this means that future profits or loss does not depend on the past.

```
df3$log_return= df3$log_price- df3$lag_price
log_return <- na.omit(df3$log_return)
acf(log_return, lag.max = 15)
```



2e, p-value < 0.05 we reject reject the null hypotesis, hence the series is correlated.

```
#Ljung-Box test
Box.test(log_return,lag=10, type = "Ljung-Box")

##
##  Box-Ljung test
##
## data:  log_return
## X-squared = 26.157, df = 10, p-value = 0.003534
```

3

$$V_t = \varphi_o + \varphi_{1vt-1} + a_t$$

$$\mu_{vt} = \frac{\varphi_o}{1 - \varphi_1}$$

$$\sigma_v^2 = \frac{\sigma_a^2}{1 - \varphi_1^2}$$

3a, The mean of the time series is 0 since φ_o = 0 .

```
#r = 0 + 0.9 lag_r +at
phi_0= 0
```

```r
mean_r = 0 / (1- .9)
mean_r

## [1] 0
```

3b, This model is stable since phi_1>0 and <1 . The acf will show an exponential decay to zero.

3c, The overall Variance of the series is 2.63

```r
####
var_r= 0.5/ (1- (.9)^2)
var_r

## [1] 2.631579
```

4

$$V_t = \theta_0 + a_t + \theta_{1at-1}$$

$$\mu_{vt} = \theta_0$$

$$\sigma_v^2 = \sigma_a^2(1 + \theta_1^2)$$

4a, Mean of the series is 5.

4b, Variance of the series is 0.031

```r
b=(-.5)*(-.5)
var_x= 0.025*(1+b)
var_x

## [1] 0.03125
```

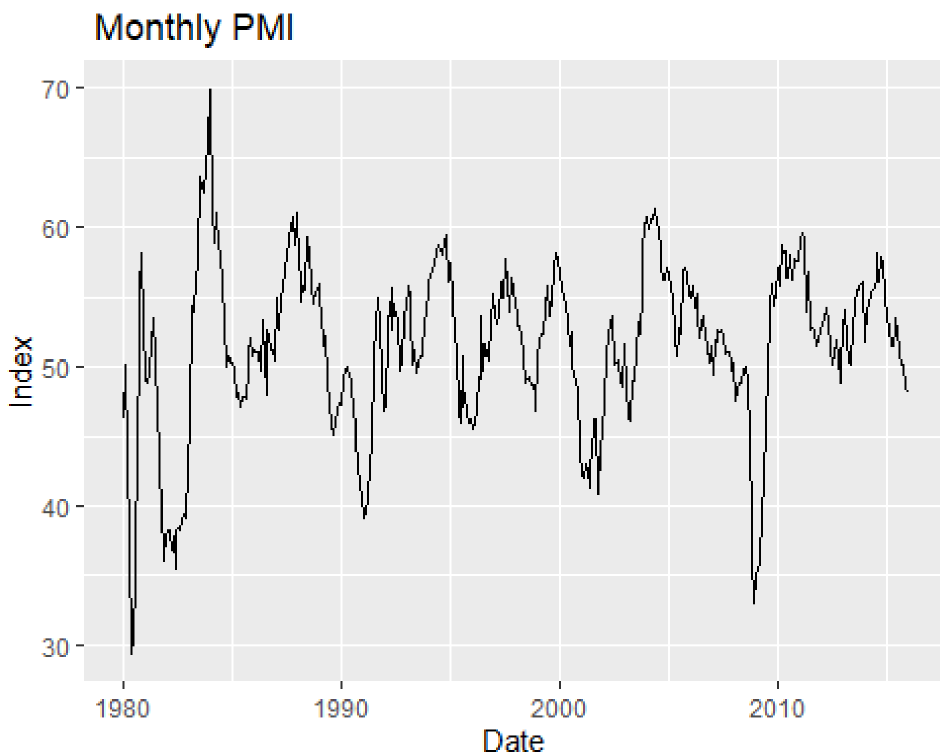
4c The MA(1) model is always stationnary beacuse the auto-correlation are 0 at lags >1.

5

5a, Import data and creat ts() from 1980.

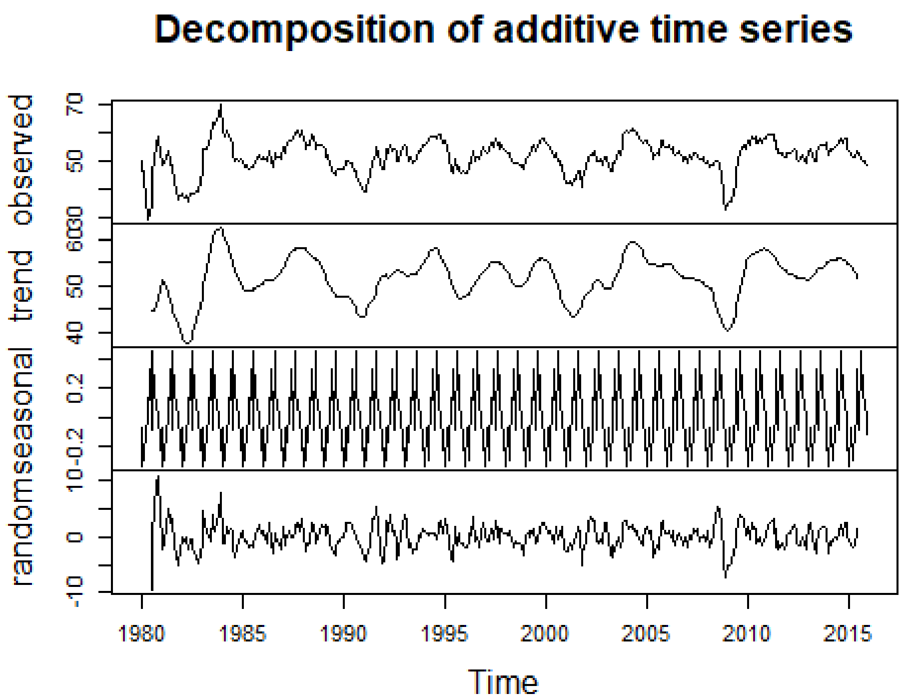
5b, Series is additive

```r
ggplot(df4,aes(x=time(df4_ts),y=df4_ts,group=1))+geom_line()+labs(title = "
Monthly PMI ")+xlab("Date")+ylab("Index")
```



5c, From the decompose graph since the scale or random is larger than the seasoninality scale, there is no seasonality in the series. There seem to be some sort of cyclic trend (up, down).

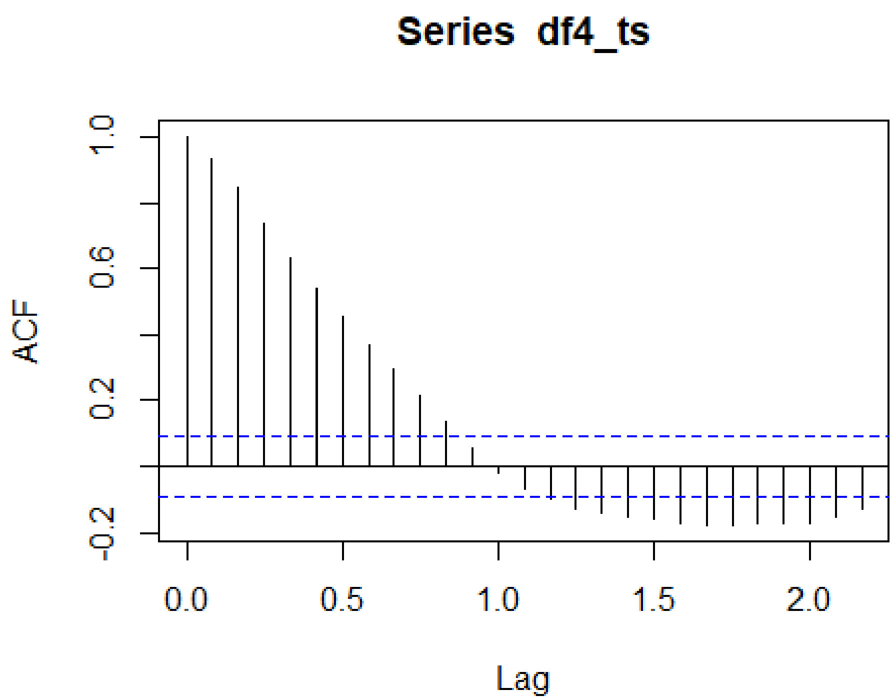
```
decom=decompose(df4_ts, type = "additive")
plot(decom)
```



5d, Analyzing serial correlation

From the ACF we suspect both a high positive and negative auto correlation of lags. The positive lags decreases rapidly to zero and contiues to negative as lag increase. The series is also stable since the ACF decays to zero quickly.

```
acf(df4_ts)
```



We reject the null hypothesis. Hence series is correlated since p-value < α.

```
#Ljung-Box test
Box.test(df4_ts,lag=10, type = "Ljung-Box")

##
##  Box-Ljung test
##
## data:  df4_ts
## X-squared = 1444.3, df = 10, p-value < 2.2e-16
```

5e, Modeling AR(2)

$$V_t = -4.53 + 1.09_{v-1} - 0.17_{v-2} + 51.4$$

ar1= phi_1 is the mean of the lag=1 , ar2 = phi_2 is the mean of the seris when lag=1.

The phi_0= -4.53 affects the mean of the overall graph this explains the change in mean over time.

Phi_1= 1 explains that or the previous 12 months would explain 100% of the present 12 month index.

phi_2= -0.16 eplains that there is a negative correlation of -16% of the past 2yrs/ 24months on the present year PMI index

```
# What if we try to model it with an AR(2) series
fit = Arima(df4_ts, order=c(2, 0, 0))
fit

## Series: df4_ts
## ARIMA(2,0,0) with non-zero mean
##
## Coefficients:
##          ar1      ar2     mean
##       1.0884  -0.1688  51.3567
## s.e.  0.0475   0.0475   1.2305
##
## sigma^2 = 4.439:  log likelihood = -934.46
## AIC=1876.93   AICc=1877.02   BIC=1893.2
```

pho_0= -4.54 ,change in mean over time.

```
#51.3567 = phi_0 / (1 - 1.0884)
phi_0 = 51.3567 * (1 - 1.0884)
phi_0

## [1] -4.539932
```

mean = 51.54

```
mean(df4_ts)

## [1] 51.54653
```

5f, From the results bellow all coefficients are significantlly differnet from zero as they pass the z test α > p-value.

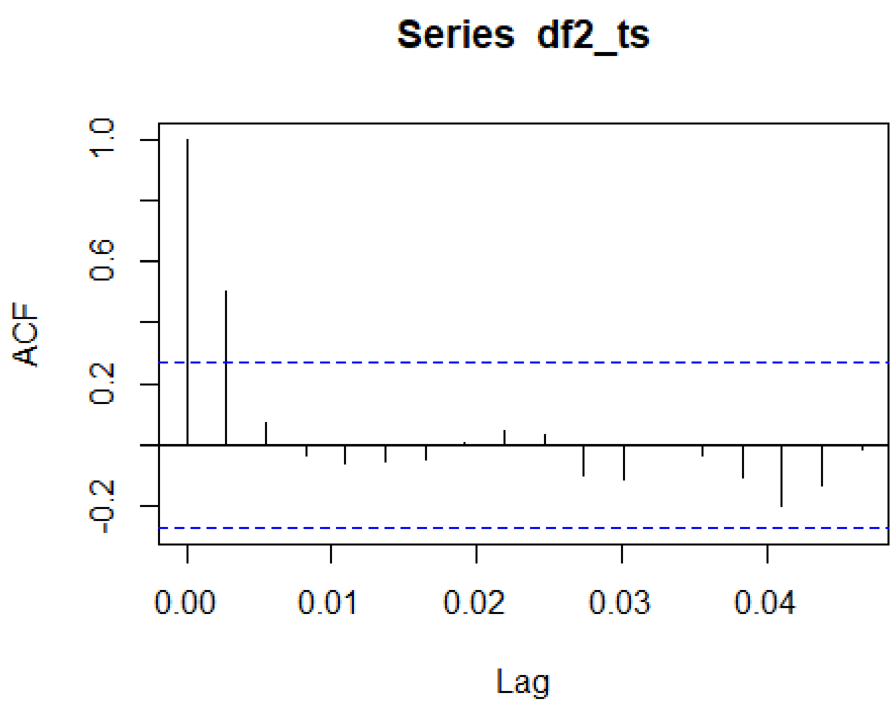
```
library(lmtest) #for coeftest
coeftest(fit)

##
## z test of coefficients:
##
##             Estimate Std. Error z value  Pr(>|z|)
## ar1         1.088378   0.047524  22.902 < 2.2e-16 ***
```

```
## ar2       -0.168805   0.047537  -3.551 0.0003837 ***
## intercept 51.356730   1.230473  41.737 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6,

6a, There seem to be a one non-zero auto correlation, hence the degree=1.

```
acf(df2_ts)
```



6b, The a_t, this is the mean shock of time (t), while theta_1 is the percentage of previous shock. This implies that there is a positive 60% increase in shock

$$V_t = 0.6_{at-1} + 219$$

```
# Try modeling this with an MA(2)
fit2 = Arima(df2_ts, order=c(0, 0,1))
fit2

## Series: df2_ts
## ARIMA(0,0,1) with non-zero mean
##
## Coefficients:
##          ma1      mean
##       0.6000  219.4313
```

```
## s.e.  0.1075    1.0399
##
## sigma^2 = 23.17:  log likelihood = -154.7
## AIC=315.4   AICc=315.9   BIC=321.26
```

6c, Both ma1 and intercept are non zero and useful since both p-values are less than 0.05.

```r
library(lmtest) #for coeftest
coeftest(fit2)
```

```
##
## z test of coefficients:
##
##             Estimate Std. Error  z value  Pr(>|z|)
## ma1          0.60002    0.10752   5.5807 2.395e-08 ***
## intercept 219.43130    1.03994 211.0045 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```