stat153_hw2_coding

1.

read in data

```r
setwd("/Users/xiaoyingliu/desktop/STAT153")
getwd()
```

```
## [1] "/Users/xiaoyingliu/Desktop/STAT153"
```

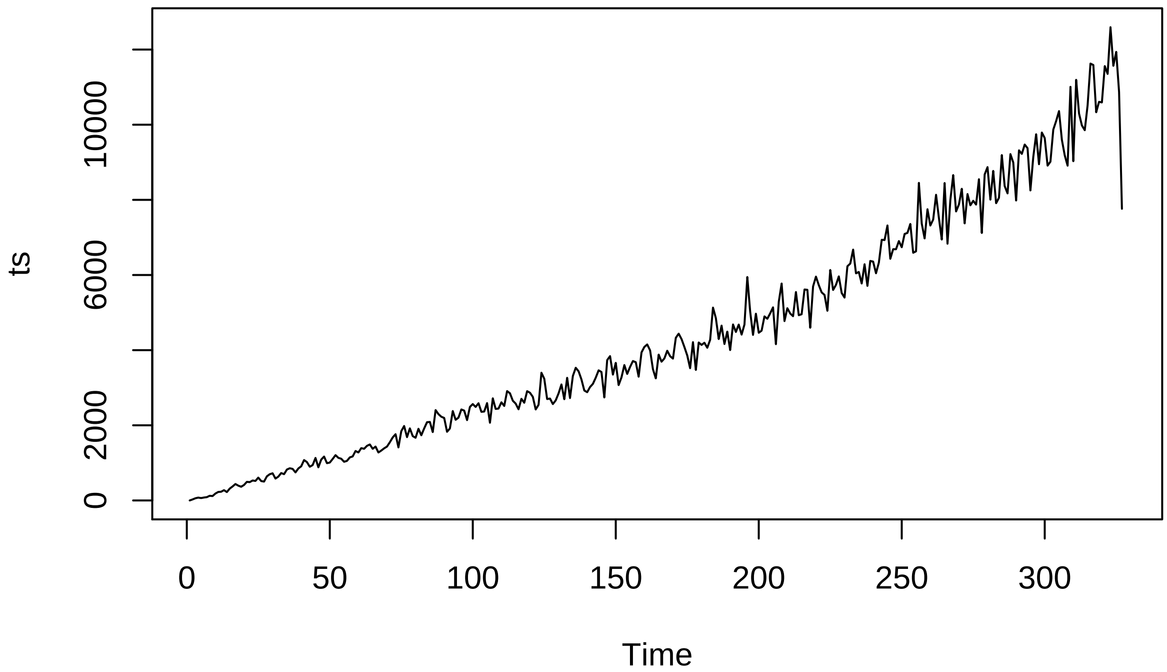
```r
?read.csv

df=read.csv("data.csv", sep = ",")
head(df)
```

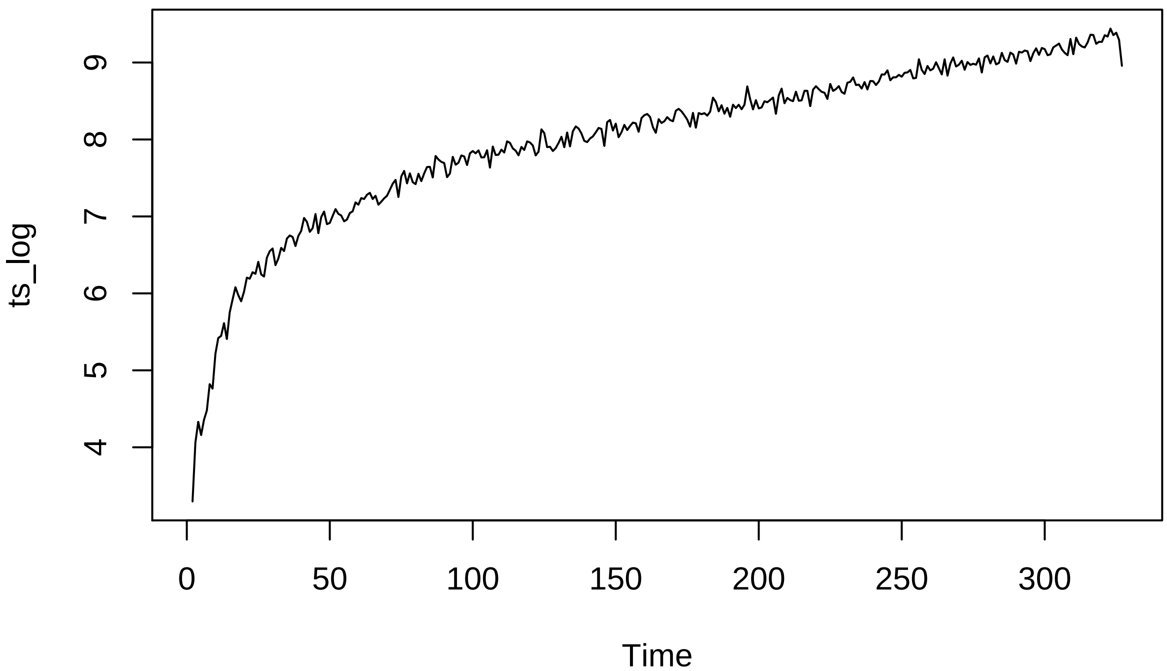
```
##   X.month submissions historical_delta
## 1  Jul-91           0               -2
## 2  Aug-91          27               -1
## 3  Sep-91          58                0
## 4  Oct-91          76                0
## 5  Nov-91          64                0
## 6  Dec-91          78                0
```

(a)stabalize variance

```r
#only take submissions and transform into a time series data
ts=as.ts(df[,2])
plot(ts)
```




```r
#I would expect that the submission have a non-constant variance.
ts_log=log(ts)
plot(ts_log)
```

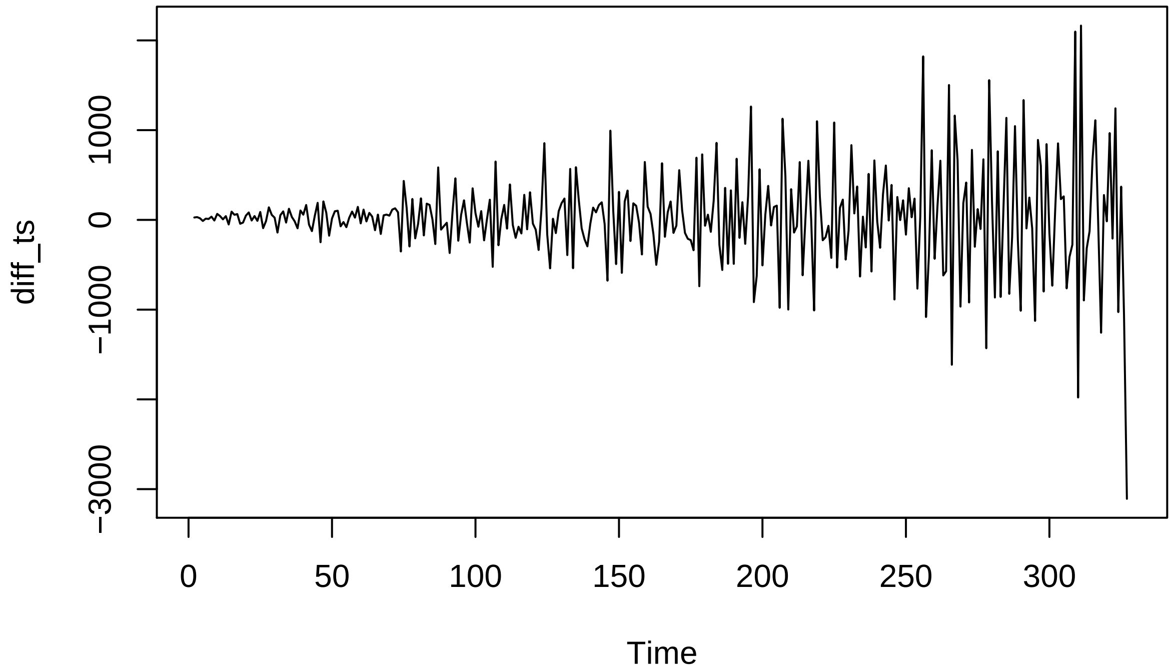



```r
#from the log plot, the variance are constant with time.
```

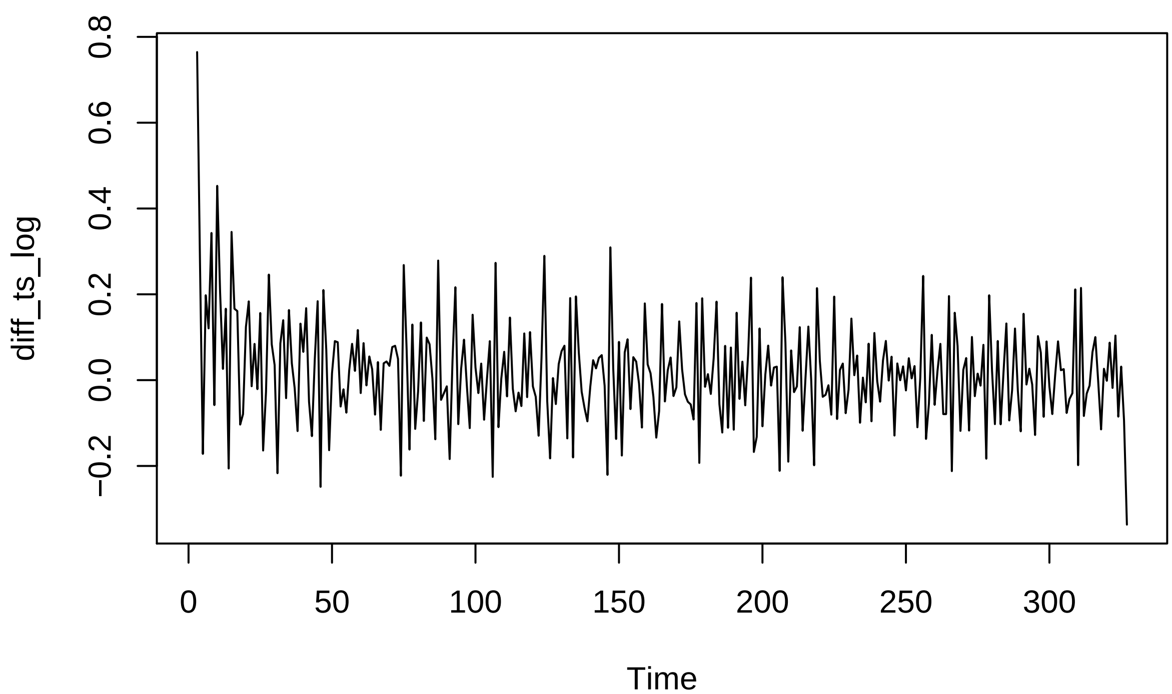
(b)difference both data

```r
diff_ts=diff(ts,differences=1,lag=1)

plot(diff_ts)
```




```r
diff_ts_log=diff(ts_log,differences=1,lag=1)
plot(diff_ts_log)
```



```
#log plot looks like white noise. However, the original data does not look like white noise at all.
```

(c)predictions for submissions in september

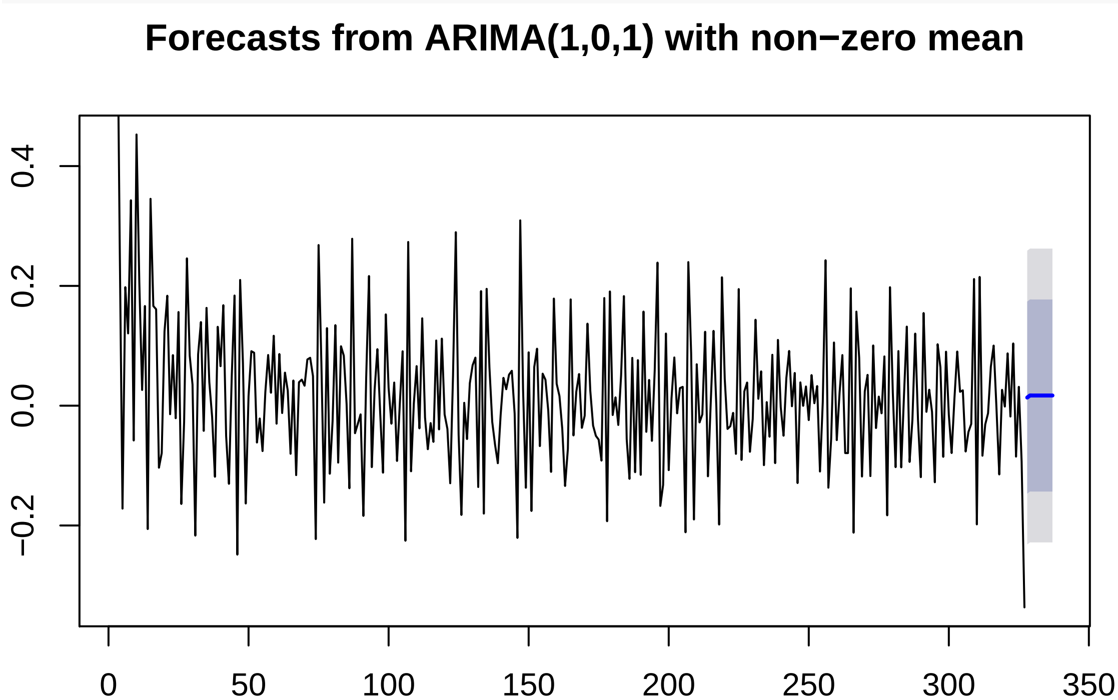
```
#fit a arima model. Since ts_log is a stationary time series, we apply arima.
library(zoo)
```

```
## Warning: package 'zoo' was built under R version 3.4.4
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
diff_ts_log=as.ts(as.zoo(diff_ts_log)[-(1)])
temp_diff=forecast(Arima(y = diff_ts_log,order = c(1,0,1)))
step=temp_diff$mean[1]
last_value=ts_log[326]
prediction=exp(last_value+step)
#diff_ts_log[324] shows ts[326],why?
prediction
```

```
## [1] 11018.46
```

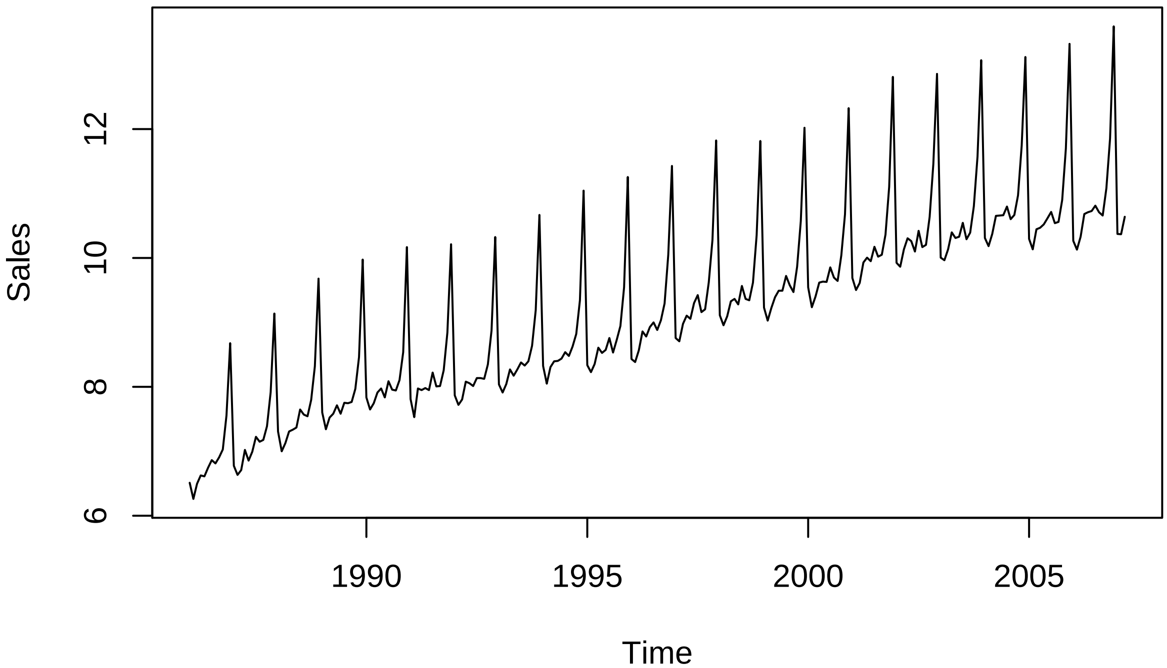
```
plot( forecast(Arima(y = diff_ts_log,order = c(1,0,1))) )
```



2

(a)

```
#load time series retail
data(retail)
#squre root plot of retail data
yt=sqrt(retail)
plot(yt)
```




```
#from the plot, we can see that there is trend as well as seasonality. The submission is increasing for
```

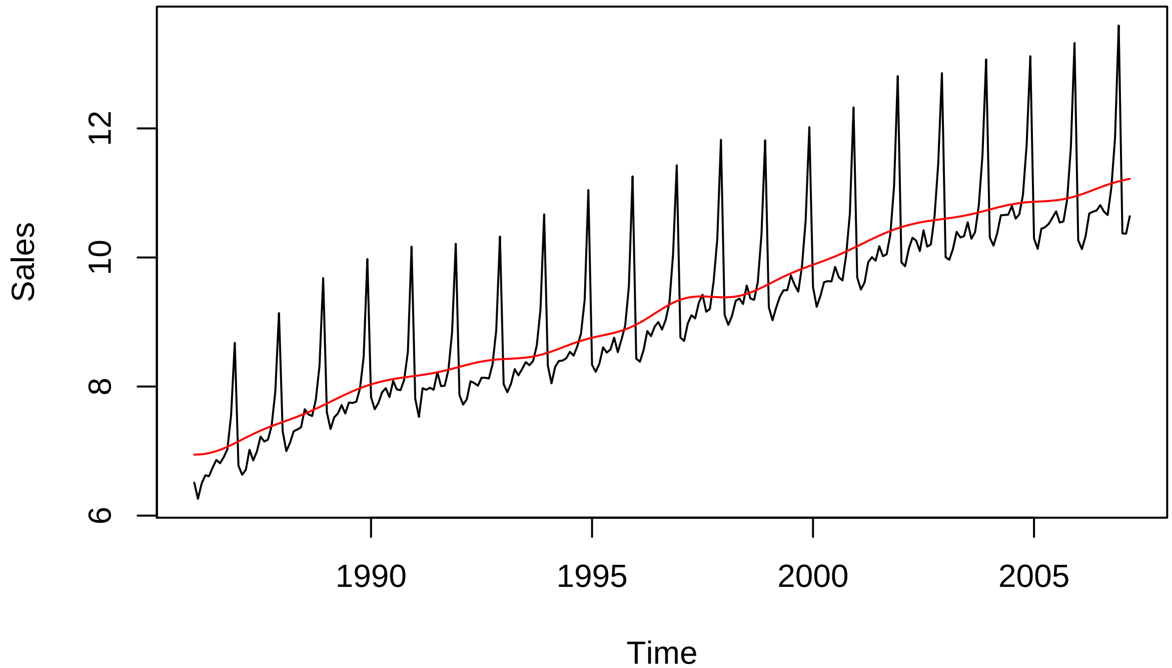
(b)

```
head(retail)
```

```
##       Jan  Feb  Mar  Apr  May  Jun
## 1986 42.4 39.2 42.2 43.9 43.7 45.5
```

```
t=as.numeric(time(yt))
x1=t
x2=cos(pi*t/6)
x3=sin(pi*t/6)
x4=cos(pi*t/3)
x5=sin(pi*t/3)
x6=cos(pi*t/2)
x7=sin(pi*t/2)
x8=cos(2*pi*t/3)
x9=sin(2*pi*t/3)
x10=cos(5*pi*t/6)
```

```r
x11=sin(5*pi*t/6)
x12=cos(pi*t)
fit1=lm(yt ~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12)
plot(yt)
lines(t, fit1$fitted.values, col='red')
```



(c)

```r
length(retail)
```

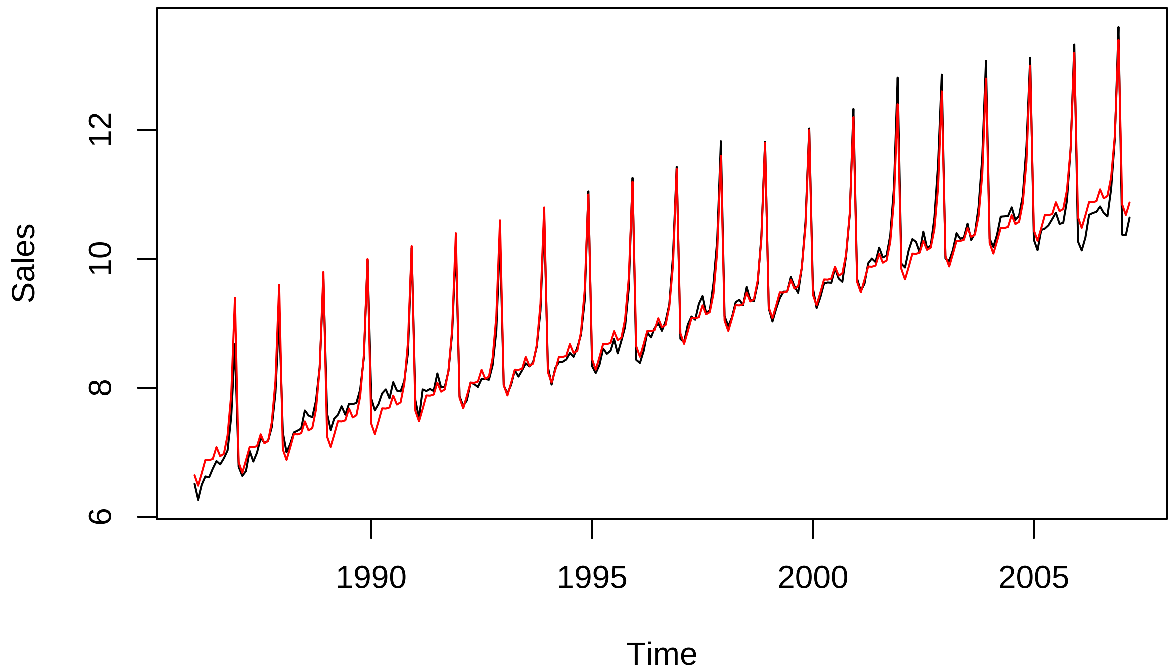
```
## [1] 255
```

```r
generator=function(x){
  v=rep(0,255)
  for(i in 255){
     v[seq(x,255,by=12)]=1
  }
  return(v)
  }


x1=t
x2=generator(1)
x3=generator(2)
x4=generator(3)
x5=generator(4)
x6=generator(5)
x7=generator(6)
x8=generator(7)
x9=generator(8)
```

```
x10=generator(9)
x11=generator(10)
x12=generator(11)

fit2=lm(yt ~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12)
plot(yt)
lines(t, fit2$fitted.values, col='red')
```



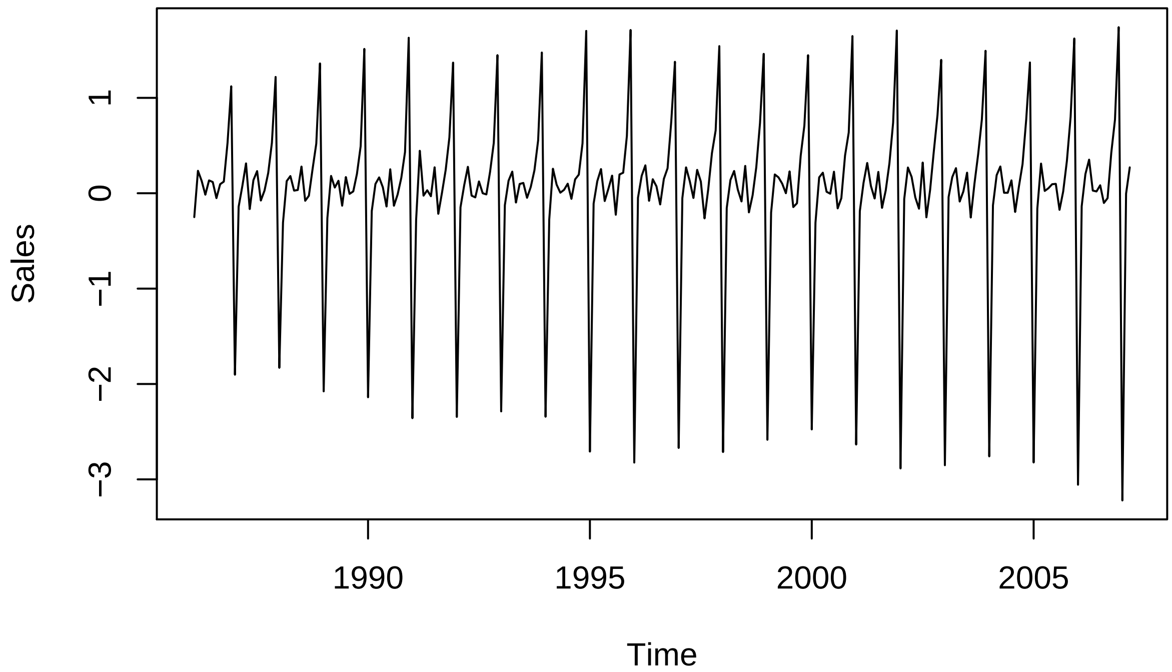
(d)comparison

the fitted values are quite different,first model captures the trend of data. teh second model cpature seasonality of data.

(e)

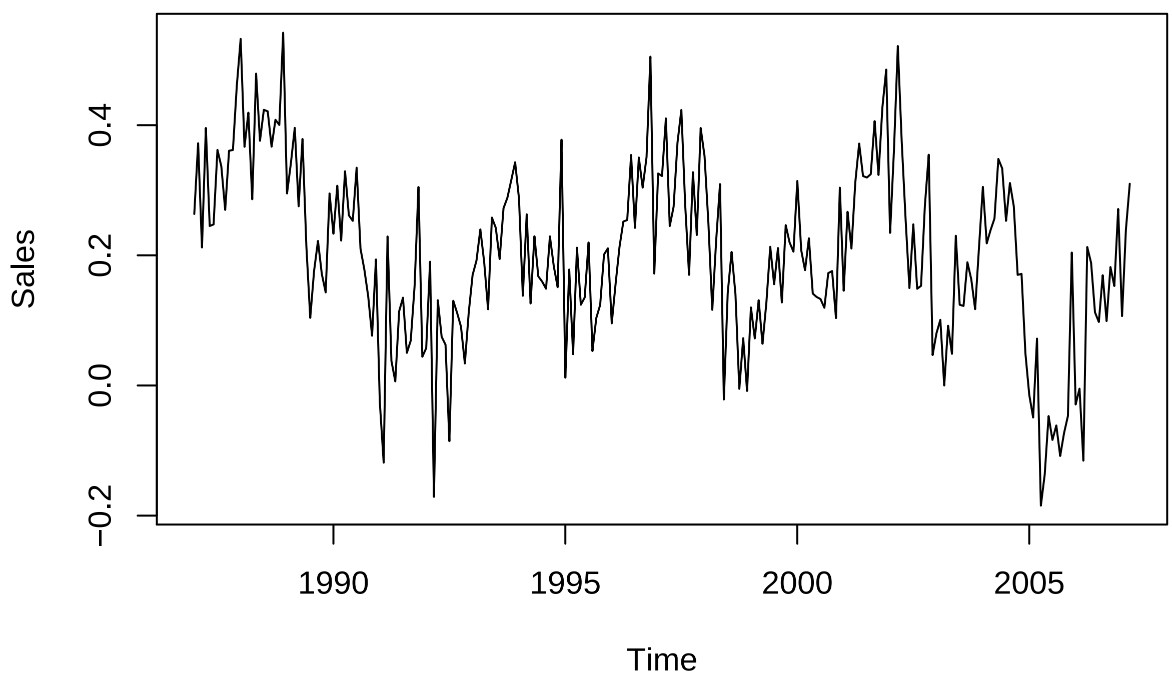
i.

```
delta_yt=diff(yt)
plot(delta_yt)
```



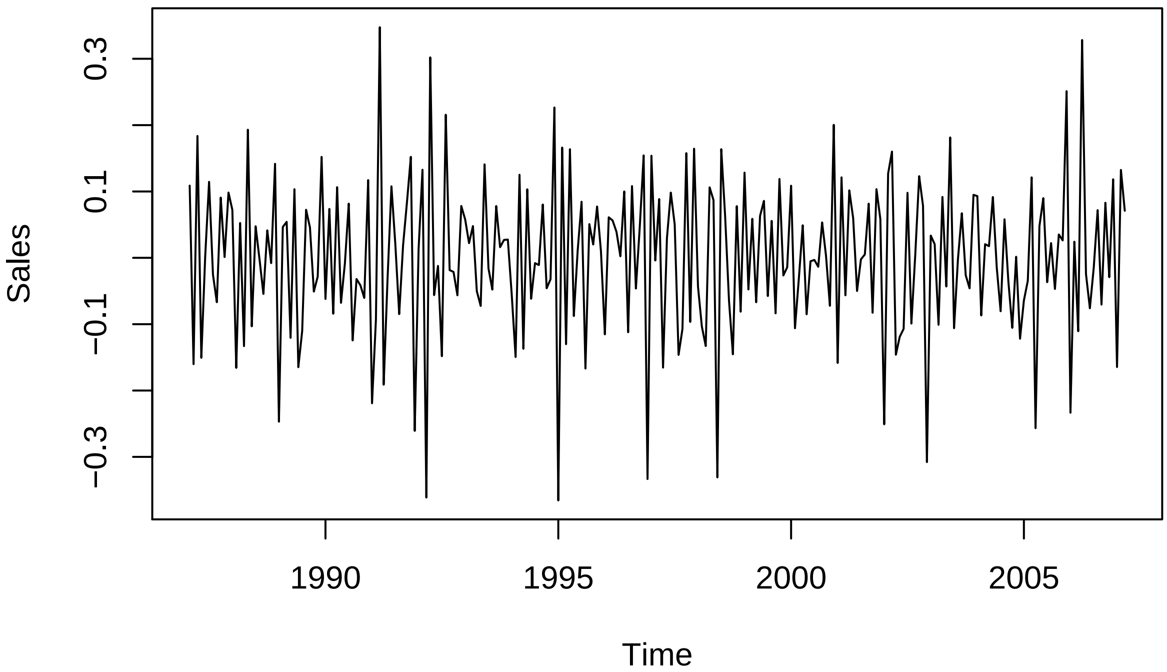
#ii.

```
delta12_yt=diff(yt,lag=12)
plot(delta12_yt)
```



iii.

```
delta_delta12_yt=diff(delta12_yt,lag=1)
plot(delta_delta12_yt)
```



#the last one looks like white noise.