Trend estimation

Dawu Liu

1. Least squares estimation

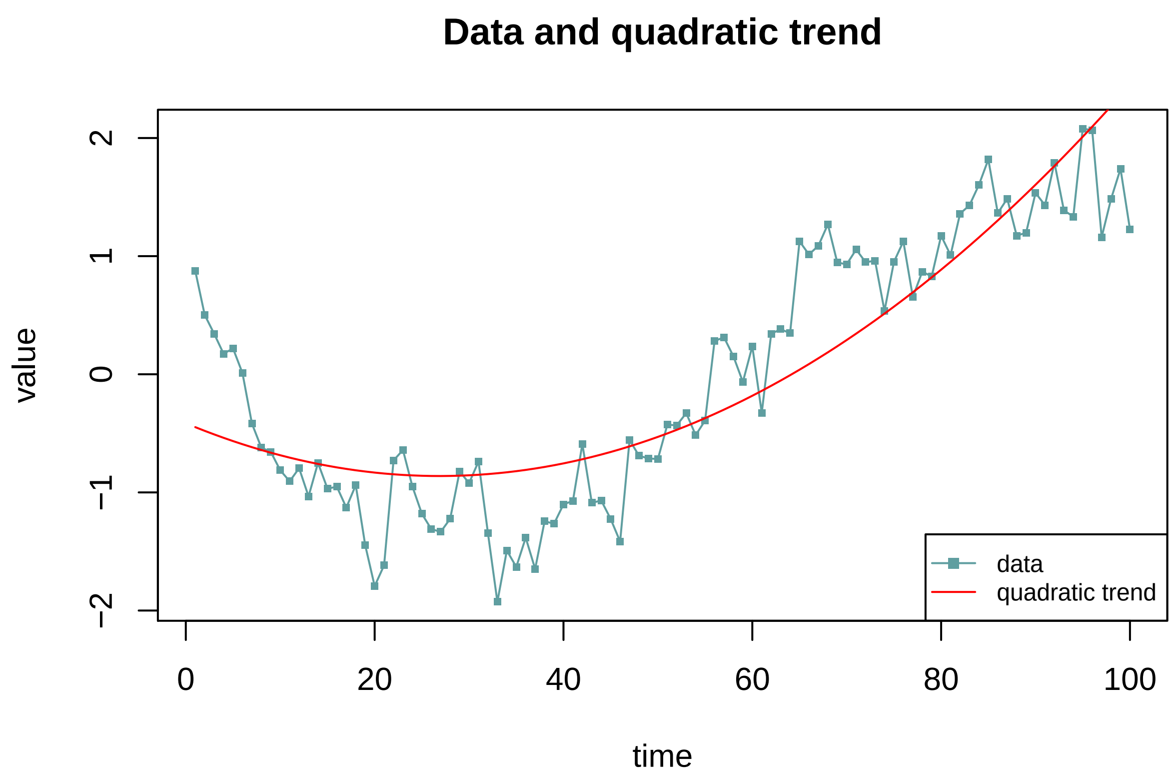
(a) Compute a least squares estimate for a quadratic trend model.

i) Give the coefficients.

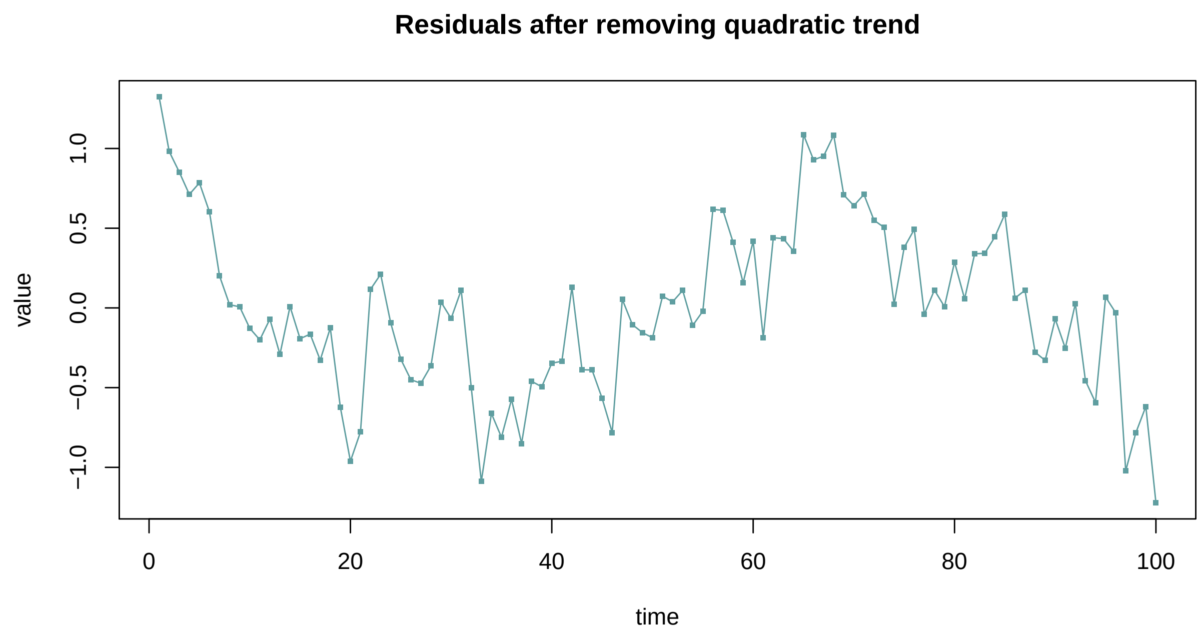
```
##    (Intercept)             t        I(t^2)
## -0.4147689817 -0.0332342602  0.0006186195
```

The coefficients are: -0.0332342602 for t, 0.0006186195 for t^2.

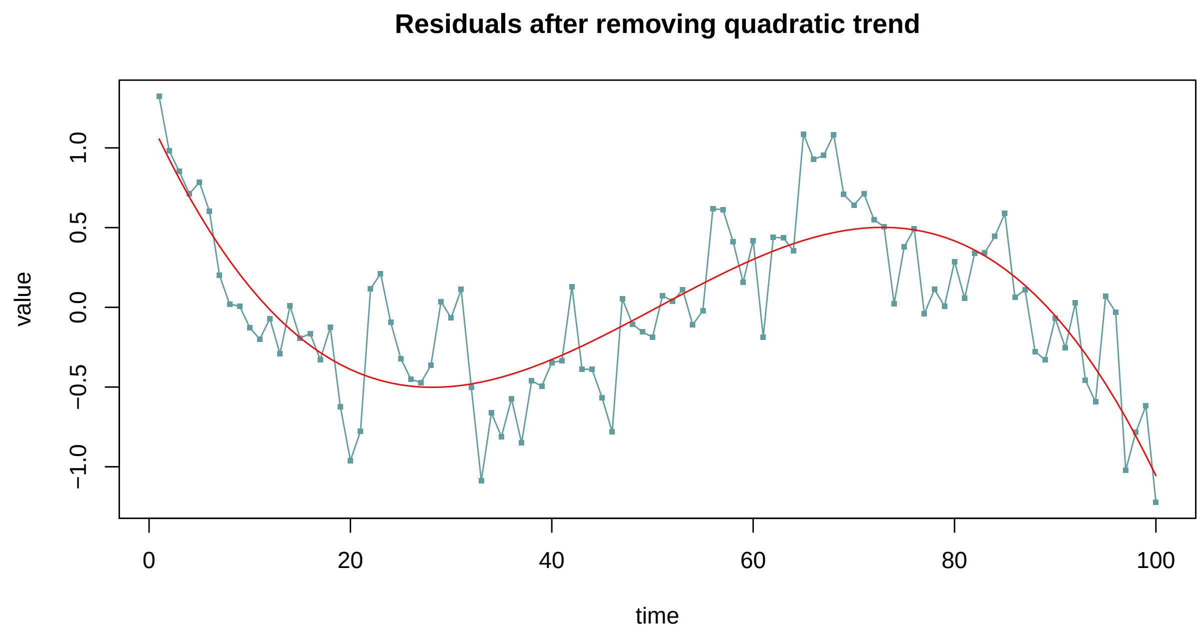
ii) Plot the data together with the estimated trend model.



iii) Plot the residuals obtained by subtracting the model from the data.



iv) Do the residuals show trend?



We can easily fit a 3rd degree polynomial trend into the residuals, just like the graph above. Yes, it's pretty clear that the residuals have a 3rd degree polynomial trend. Goes down, then up, then down again in a curve.

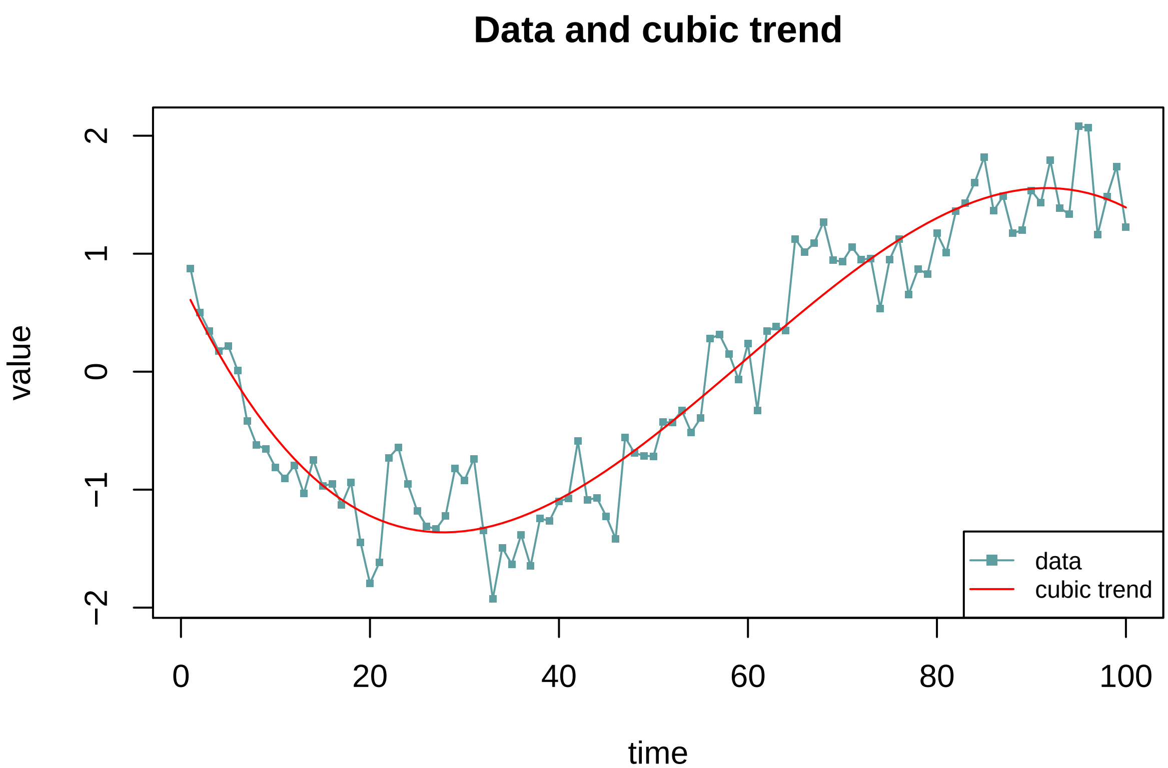
(b) Compute a least squares estimate for a cubic trend model.

i) Give the coefficients.

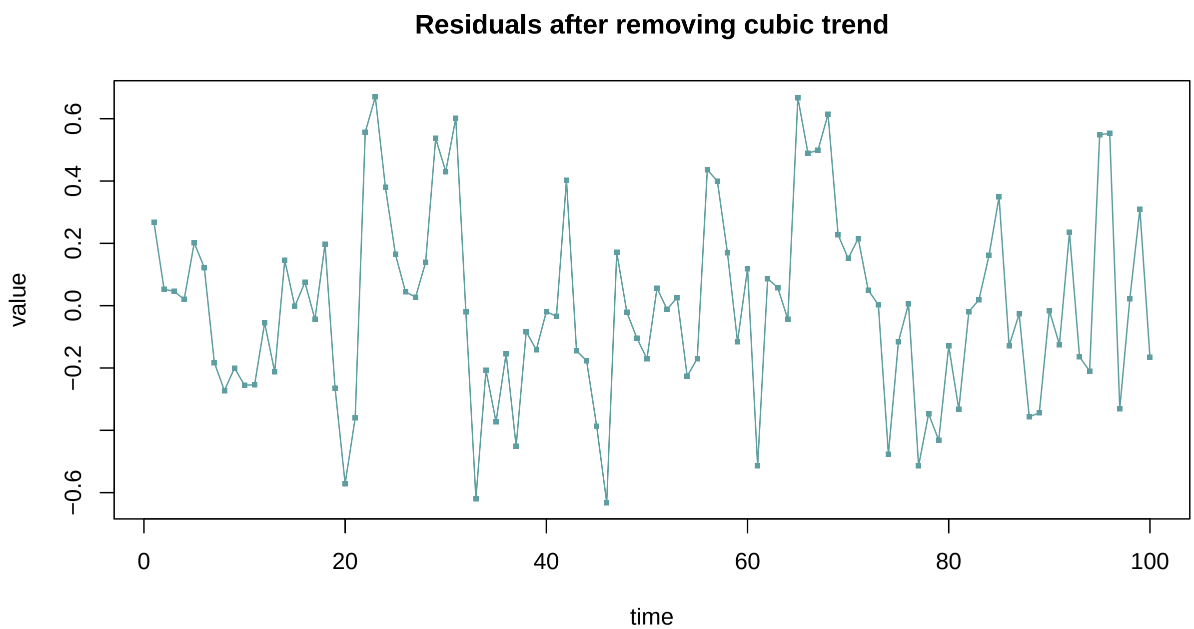
```
##    (Intercept)             t         I(t^2)        I(t^3)
##   0.7760731619 -0.1712975472  0.0040190825 -0.0000224453
```

The coefficients are: –0.1712975472 for t, 0.0040190825 for t^2, -0.0000224453 for t^3.

ii) Plot the data together with the estimated trend model.

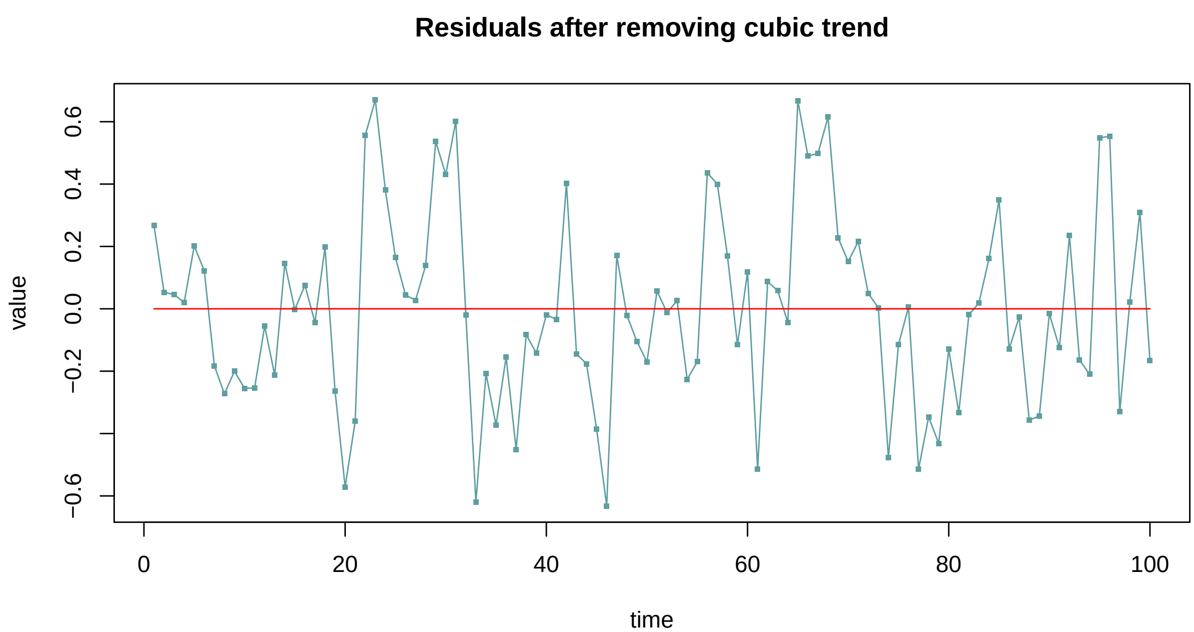


iii) Plot the residuals obtained by subtracting the model from the data.



iv) Do the residuals show trend?

To make it easy to check, I fit a quadratic model to the time series.



There doesn't appear to be any trend in the residuals. The graphs above is what happens after trying to fit an quadratic model to the residuals. And it appears to be a horizontal line.

(c) Which trend model is better, quadratic or cubic?

Cubic model is better. By subtracting the model from the data, cubic model residuals don't appear to have any trend, but quadratic model residuals still have a clear trend.

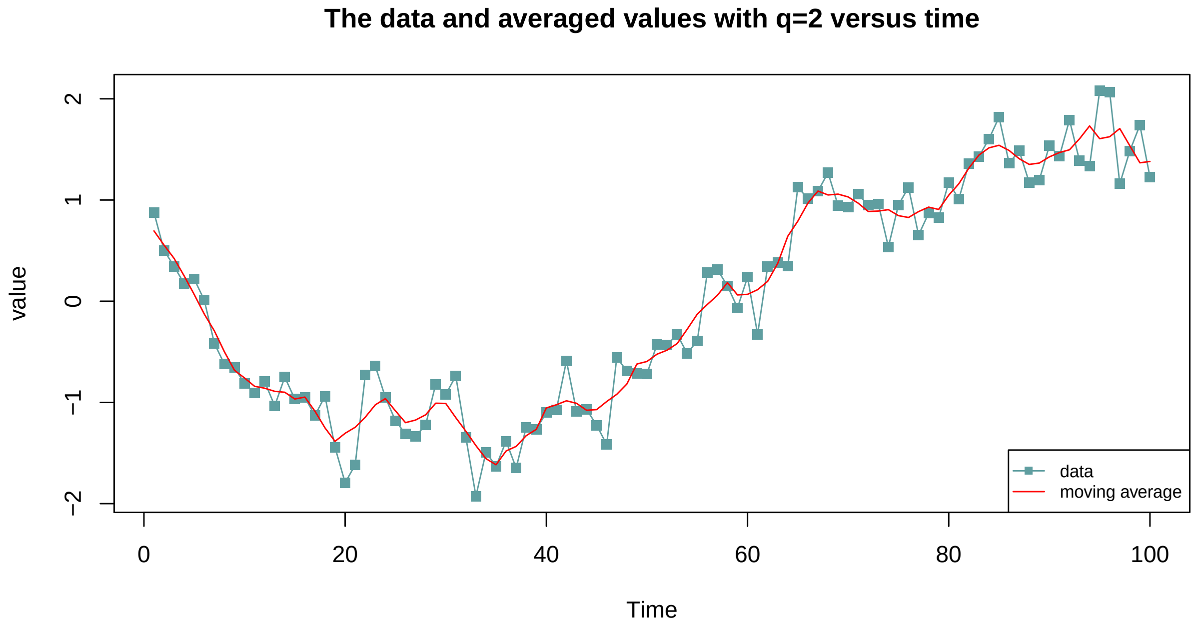
Code used:

```
library(forecast)
data = read.delim("E:/2021 Fall/STAT 485/data/project4_data.txt", header = FALSE)

# (a)
# i)
t = 1:length(data$V1)
quadratic = lm(data$V1 ~ t + I(t^2))
quadratic$coefficients
# ii)
plot(data$V1, type="o", pch=15, cex=0.5, col="cadetblue", xlab="time", ylab="value",
     main="Data and quadratic trend")
lines(fitted(quadratic), col="red",type="l")
legend("bottomright", legend = c("data", "quadratic trend"),cex = 0.75,
       pch = c(15,NA), lty= c(1,1), col = c("cadetblue", "red"))
# iii)
plot(quadratic$residuals, type="o", col="cadetblue", pch=15, cex=0.5, xlab="time", ylab="value",
     main="Residuals after removing quadratic trend")
# iv)
plot(quadratic$residuals, type="o", col="cadetblue", pch=15, cex=0.5, xlab="time", ylab="value",
     main="Residuals after removing quadratic trend")
lines(fitted(lm(quadratic$residuals ~ t + I(t^2) + I(t^3))), col="red", type="l")

# (b)
# i)
t = 1:length(data$V1)
cubic = lm(data$V1 ~ t + I(t^2) + I(t^3))
cubic$coefficients
# ii)
plot(data$V1, type="o", pch=15, cex=0.5, col="cadetblue", xlab="time", ylab="value",
     main="Data and cubic trend")
lines(fitted(cubic), col="red",type="l")
legend("bottomright", legend = c("data", "cubic trend"),cex = 0.75,
       pch = c(15,NA), lty= c(1,1), col = c("cadetblue", "red"))
# iii)
plot(cubic$residuals, type="o", col="cadetblue", pch=15, cex=0.5, xlab="time", ylab="value",
     main="Residuals after removing cubic trend")
# iv)
plot(cubic$residuals, type="o", col="cadetblue", pch=15, cex=0.5, xlab="time", ylab="value",
     main="Residuals after removing cubic trend")
lines(fitted(lm(quadratic$residuals ~ t + I(t^2))), col="red", type="l")
```

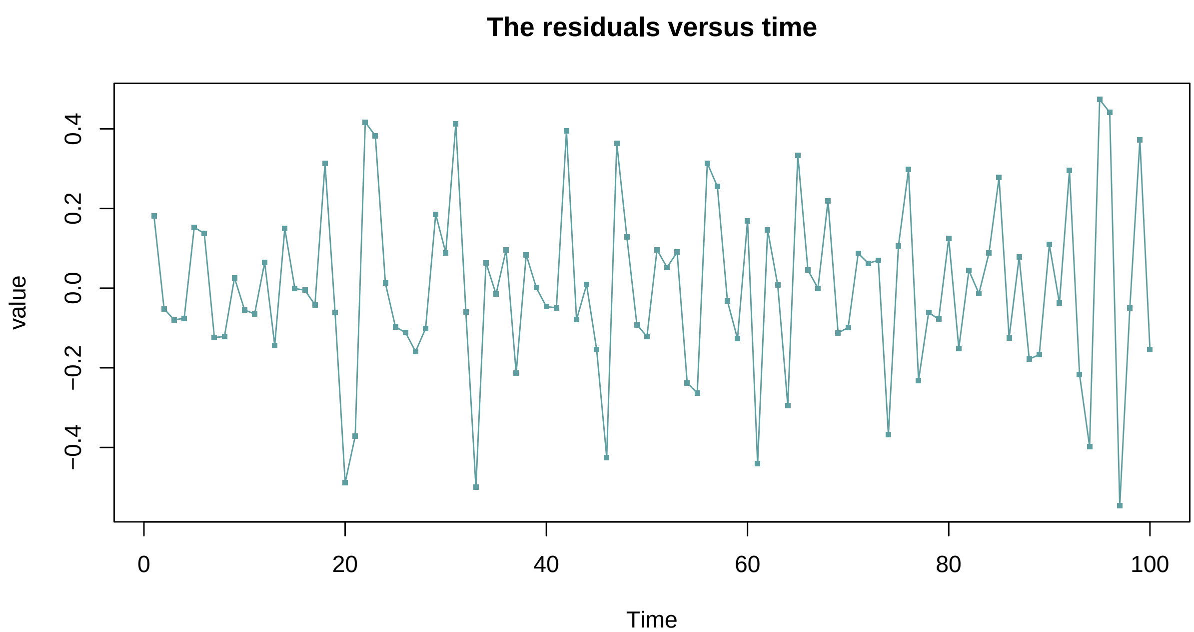
2. Moving average

Apply a moving average with q = 2 and equal weights (Example 4.2.2.1) to the data.

(a) Plot the moving average and the original data on the same plot.

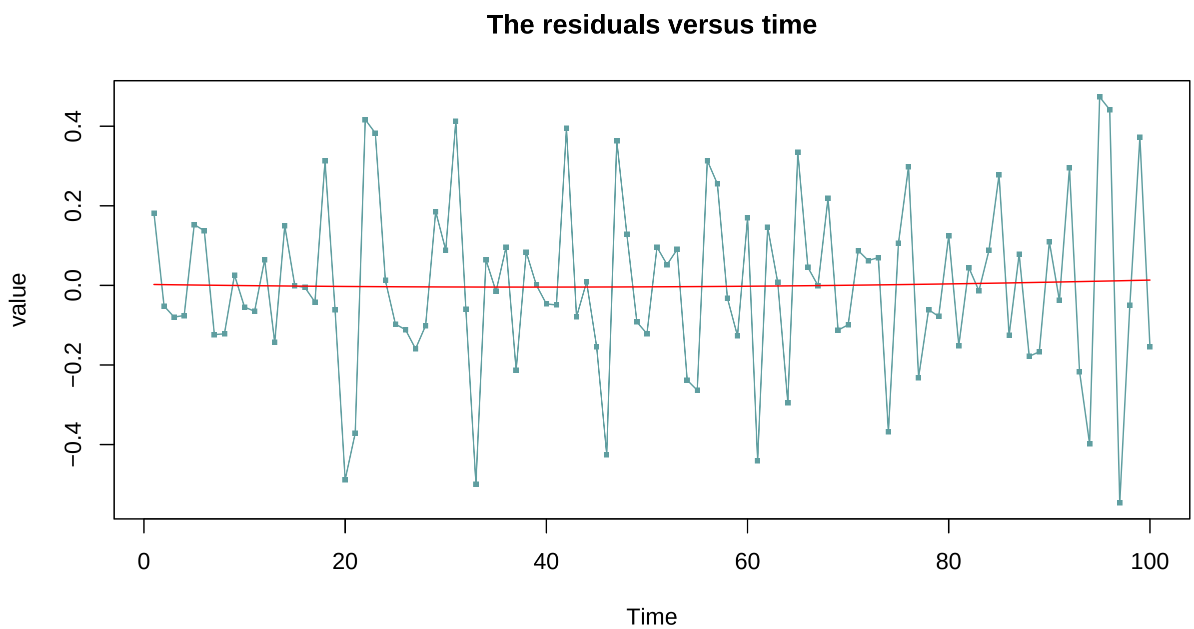


(b) Plot the residuals obtained by subtracting the series obtained by the moving average from the original data.



(c) Do the residuals show trend?

To make it easy to check, I fit a quadratic model to the time series.



There doesn't appear to be any trend in the residuals. The graphs above is what happens after trying to fit an quadratic model to the residuals. And it appears to be pretty much a horizontal line.

Code used:

```
# (a)
data = read.delim("E:/2021 Fall/STAT 485/data/project4_data.txt", header = FALSE)
data.ts = ts(data$V1)

#Assume the value at t_-1 and t_0 the same as t_1, the value at t_101 and t_102 the same as t_100,
data.ts.new = c(data.ts[1], data.ts[1], data.ts, data.ts[length(data.ts)], data.ts[length(data.ts)])
data.ma = ma(data.ts.new, order = 5) #order= 2q+1, q=2 -> 2q+1 = 5 (2 past, 1 current, 2 next)
data.ma = data.ma[is.na(data.ma)==FALSE] # remove NA values at t_-1, t_0, t_101, t_101

#plot the data and the averaged values for q = 2.
plot.ts(data.ts, type="o", pch=15, col="cadetblue",
        ylab="value", main="The data and averaged values with q=2 versus time")
lines(data.ma, col="red") #moving average
legend("bottomright", legend = c("data", "moving average"),cex = 0.75,
       pch = c(15,NA), lty= c(1,1), col = c("cadetblue", "red"))

# (b)
data.res = data.ts - data.ma # residuals
plot.ts(data.res, type="o", col="cadetblue", pch=15, cex=0.5,
        ylab="value", main="The residuals versus time")

# (c)
plot.ts(data.res, type="o", col="cadetblue", pch=15, cex=0.5,
        ylab="value", main="The residuals versus time")\
# try to fit a quadratic model to it and check if there's any trend
lines(fitted(lm(data.res ~ t + I(t^2))), col="red", type="l")
```

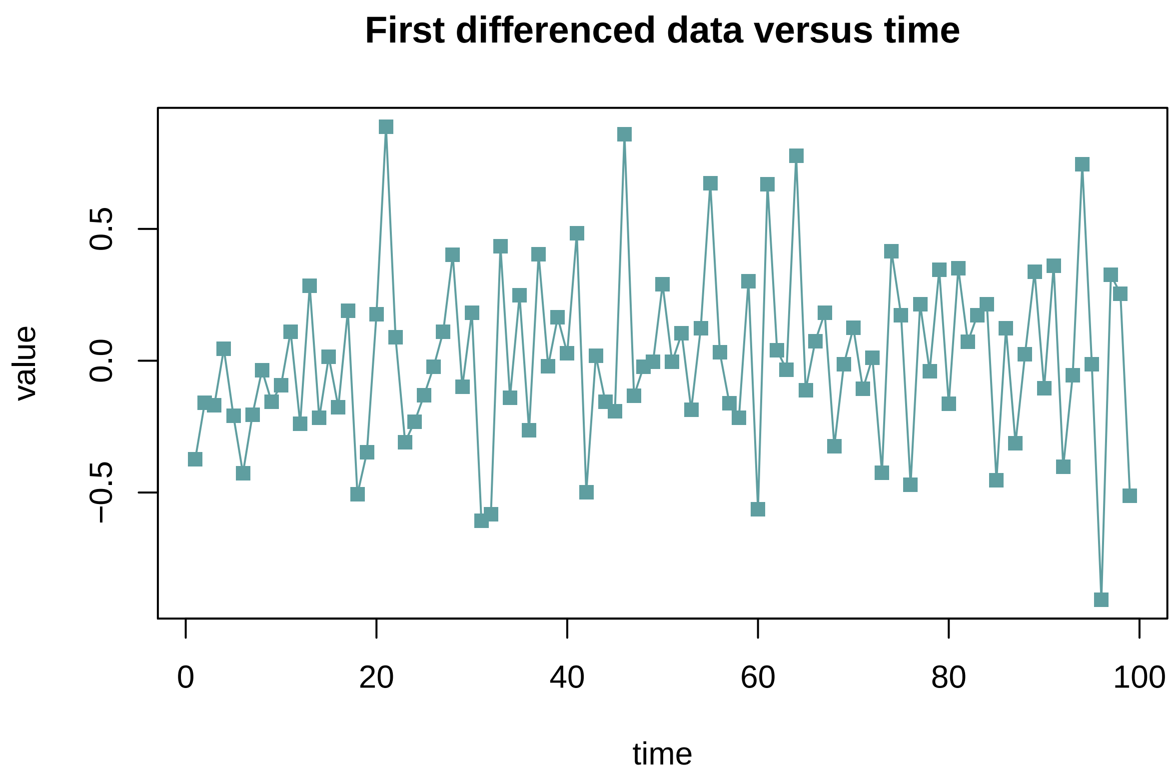
3. Differencing

(a) Apply a single difference to the data.

```
data = read.delim("E:/2021 Fall/STAT 485/data/project4_data.txt", header = FALSE)
first_diff = diff(data$V1, differences = 1)
```

i) Plot the residuals.

```
plot(first_diff, type="o", col="cadetblue", pch=15, xlab="time", ylab="value",
     main="First differenced data versus time")
```



ii) Do you think the plot of the residuals shows evidence of trend?

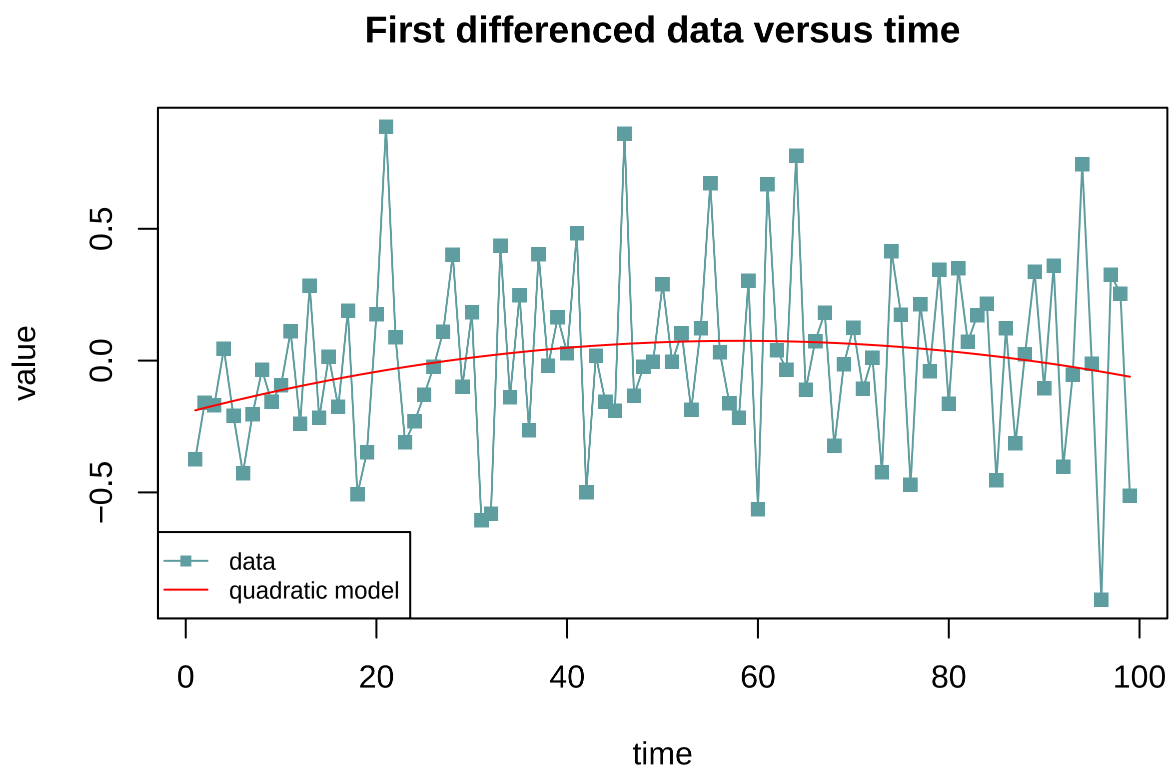
There might be a slight polynomial trend. The graph appear to go up then go down a little bit. But it's difficult too see.

iii) Fit a quadratic to the residuals using least squares, give the coefficients, and plot the fit with the residuals. Does this suggest that the residuals have trend?

```
t=1:length(first_diff)
quad = lm(first_diff ~ t + I(t^2))
quad$coefficients
```

```
##   (Intercept)             t        I(t^2)
## -0.1980299425  0.0094127425 -0.0000811257
```

```
plot(first_diff, type="o", col="cadetblue", pch=15, xlab="time", ylab="value",
     main="First differenced data versus time")
lines(fitted(quad), col="red",type="l")
legend("bottomleft", legend = c("data", "quadratic model"),cex = 0.75,
        pch = c(15,NA), lty= c(1,1), col = c("cadetblue", "red"))
```



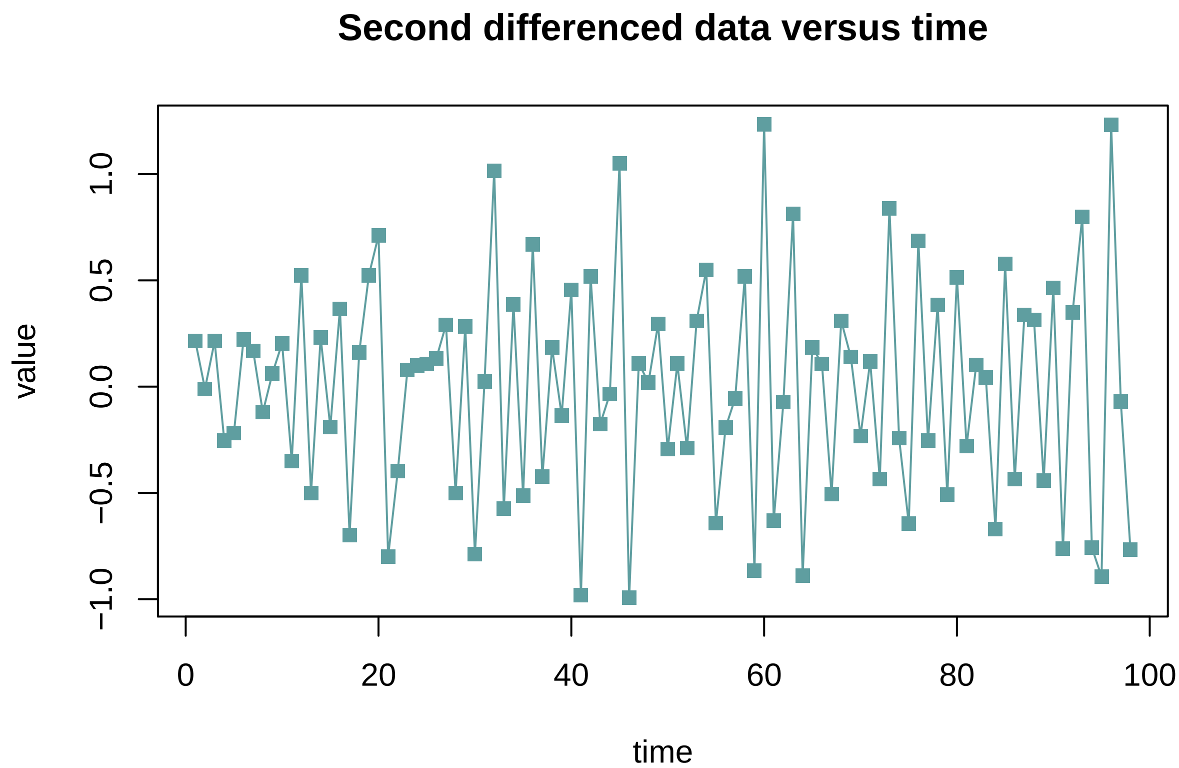
The coefficient for t, t^2 are 0.0094127425, -0.0000811257 respectively, which are sufficient large for trend. And the quadratic model fit graph also indicates there is a clear quadratic trend in the residuals.

(b) Apply a second order difference to the data.

```
second_diff = diff(data$V1, differences = 2)
```

i) Plot the residuals.

```
plot(second_diff, type="o", col="cadetblue", pch=15, xlab="time", ylab="value",
     main="Second differenced data versus time")
```



ii) Do you think the plot of the residuals shows evidence of trend?

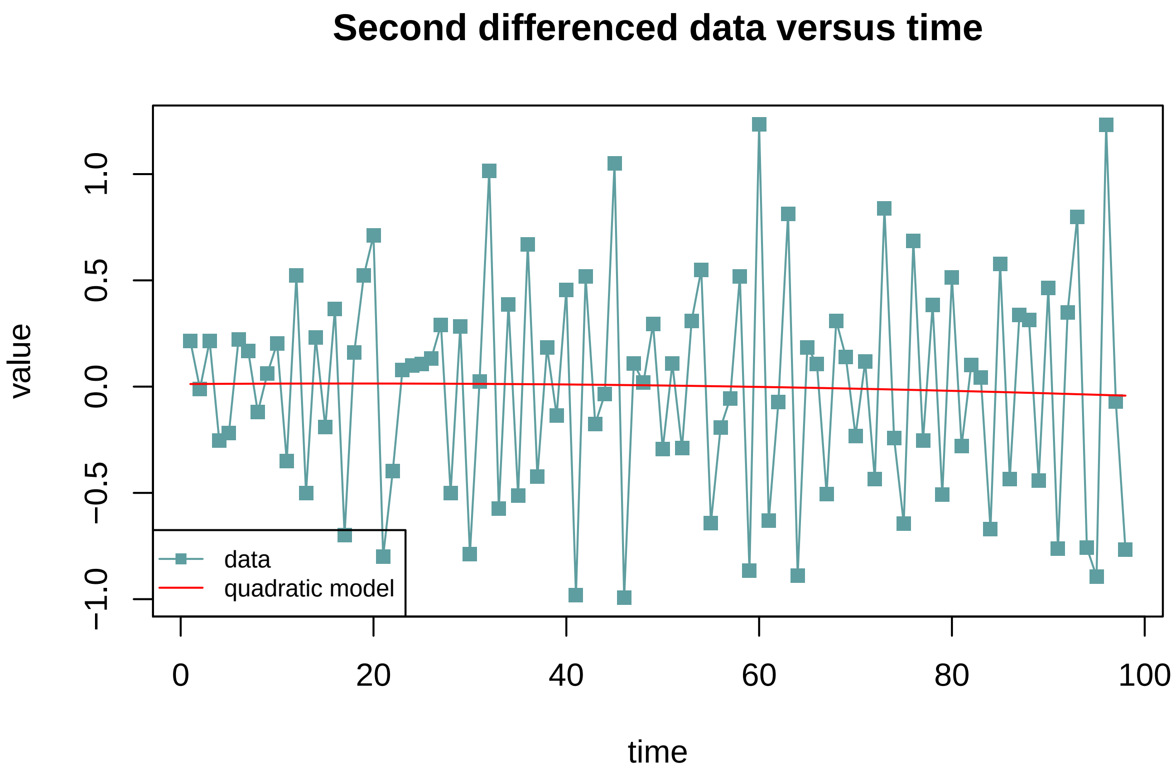
There doesn't appear to be any trend the in residuals.

iii) Fit a quadratic to the residuals using least squares, give the coefficients, and plot the fit with the residuals. Does this suggest that the residuals have trend?

```
t=1:length(second_diff)
quad2 = lm(second_diff ~ t + I(t^2))
quad2$coefficients
```

```
##   (Intercept)             t        I(t^2)
##  1.200702e-02  3.053476e-04 -8.782542e-06
```

```
plot(second_diff, type="o", col="cadetblue", pch=15, xlab="time", ylab="value",
     main="Second differenced data versus time")
lines(fitted(quad2), col="red",type="l")
legend("bottomleft", legend = c("data", "quadratic model"),cex = 0.75,
       pch = c(15,NA), lty= c(1,1), col = c("cadetblue", "red"))
```



The coefficient for t, t^2 are 3.053476e-04, -8.782542e-06 respectively, which are close to zero and much smaller than the coefficients of the first differecing. The quadratic model fit appear to be a nearly horizontal line. Indicating there is no trend in the residuals.