# STAT 443: Assignment 1

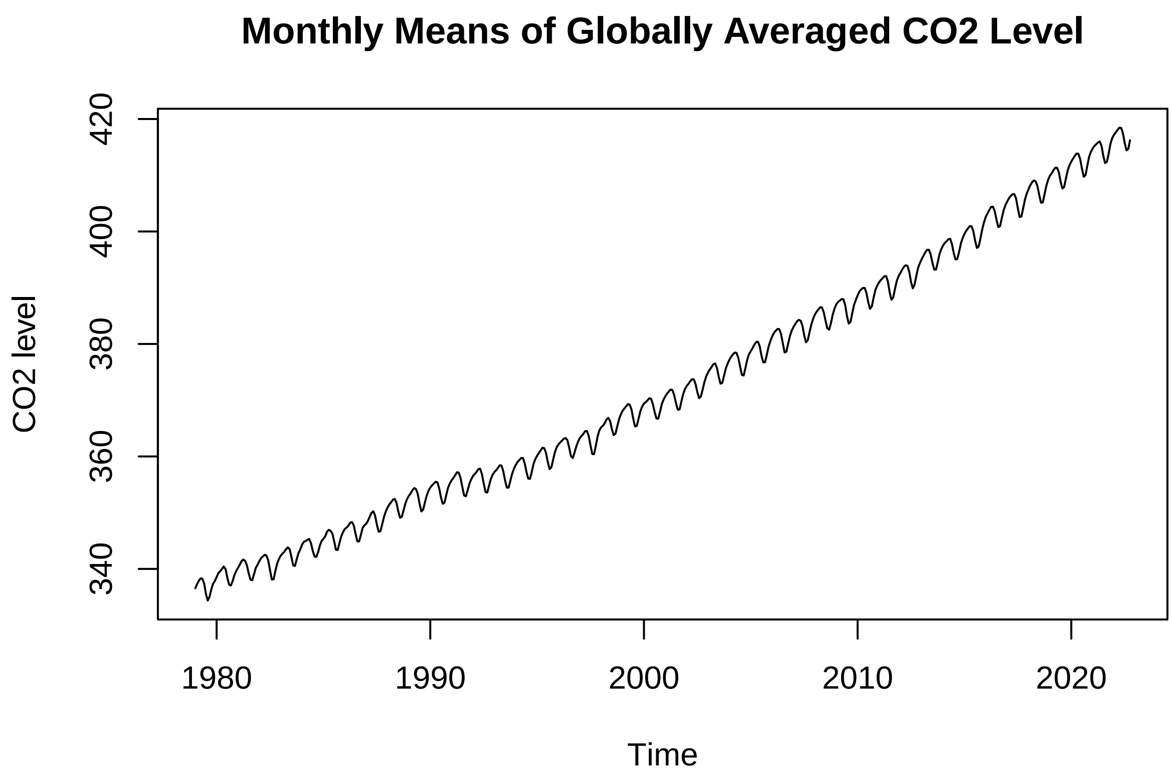
Wenxuan Zan (61336194)

2 Februray, 2022

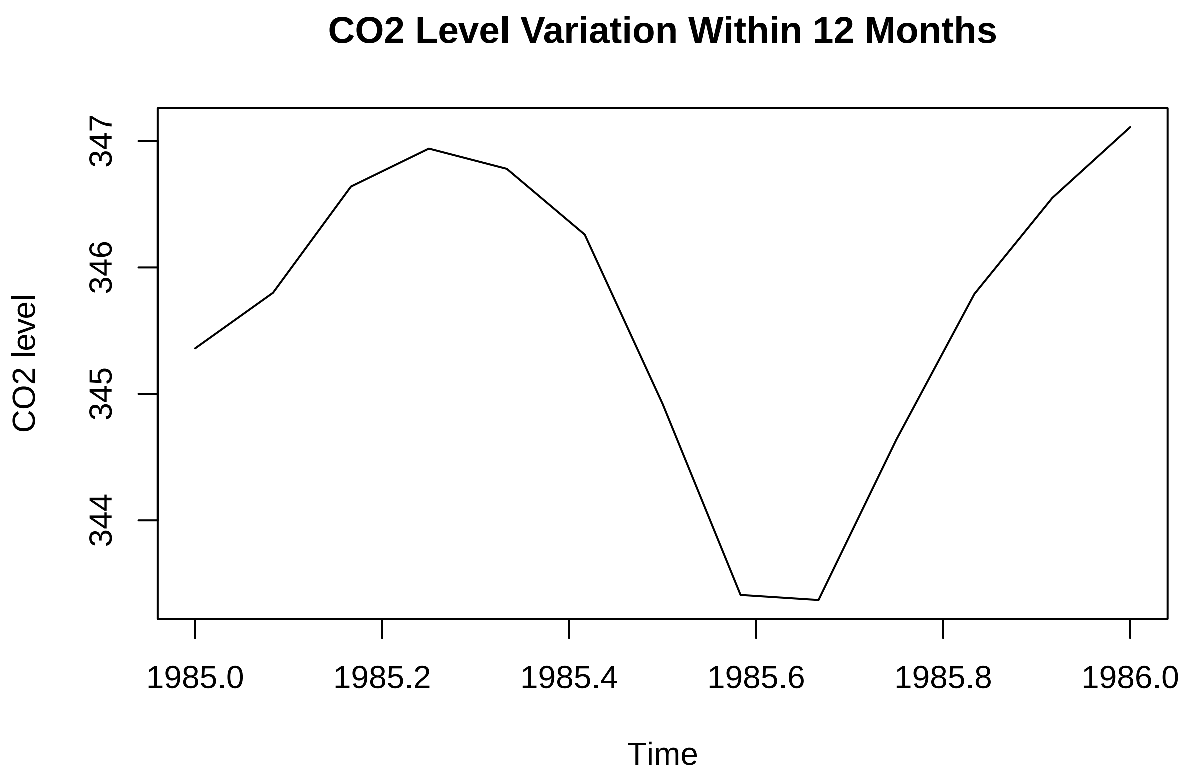
**Question 1**

a)

```
co2 <- read.csv("co2_mm_gl.csv", header = TRUE, skip = 55)
co2_ts <- ts(co2[,4], start = c(1979,1),frequency = 12)
plot(co2_ts,
     main = "Monthly Means of Globally Averaged CO2 Level",
     ylab = "CO2 level")
```




```
plot(window(co2_ts, start = c(1985,1),end = c(1986,1)),
     main = "CO2 Level Variation Within 12 Months",
     ylab = "CO2 level")
```




i) The above time series have a clear **upward trend** such that the monthly means of globally averaged CO2 level is increasing every year despite some variations.

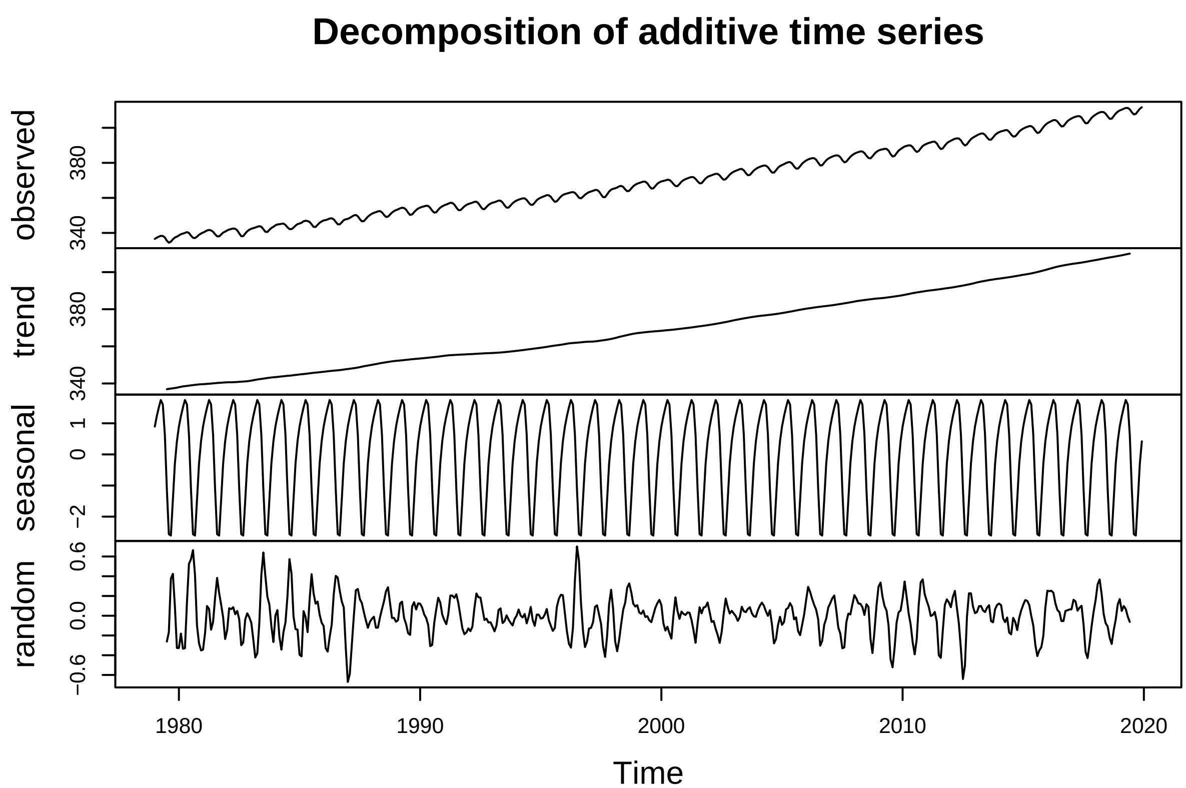
ii) There appears to be **seasonal variations** in the monthly CO2 level as when we restrict the plot to display the average CO2 level over a 12-month period, we can clear see the CO2 level is high around March and December, low around June.**An additive model is more suitable** as we can see the seasonal effect remains constant over time and the error is also constant over time, therefore an additive model would be more appropriate than a multiplicative model.

iii) **No**, the series have a clear upward increasing trend therefore it is not stationary.

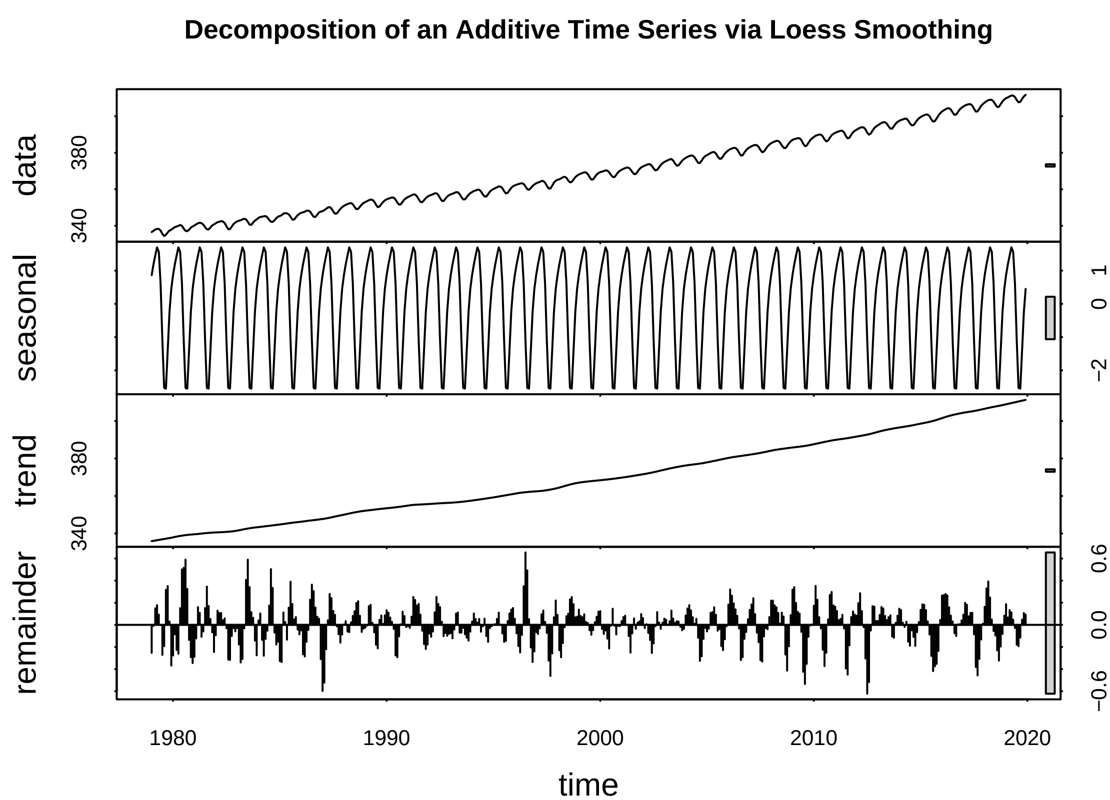
b)

```
co2_train <- window(co2_ts,start = c(1979,1),end = c(2019,12),frequency = 12)
co2_test <- window(co2_ts, start = c(2020,1), end = c(2022,10), frequency = 12)

co2_train_decom <- decompose(co2_train, type = "additive")
plot(co2_train_decom)
```




```
co2_train_loess <- stl(co2_train,s.window = "periodic")
plot(co2_train_loess,
     main = "Decomposition of an Additive Time Series via Loess Smoothing")
```



c)

```r
# MA method
ma_trend <- co2_train_decom$trend
# Loess smoothing
loess_trend <- co2_train_loess$time.series[,"trend"]
# Fitted Models
lm_ma <- lm(co2_train ~ ma_trend)
lm_loess <- lm(co2_train ~ loess_trend)
summary(lm_ma)
```

```
##
## Call:
## lm(formula = co2_train ~ ma_trend)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1127 -1.2961  0.4904  1.3657  1.9759
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.581746   1.250670  -0.465    0.642
## ma_trend     1.001573   0.003379 296.440   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.533 on 478 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.9946, Adjusted R-squared:  0.9946
## F-statistic: 8.788e+04 on 1 and 478 DF,  p-value: < 2.2e-16
```

```r
summary(lm_loess)
```

```
##
## Call:
## lm(formula = co2_train ~ loess_trend)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.045 -1.317  0.527  1.373  2.001
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.38778    1.19993   0.323    0.747
## loess_trend  0.99895    0.00324 308.272   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.529 on 490 degrees of freedom
## Multiple R-squared:  0.9949, Adjusted R-squared:  0.9949
## F-statistic: 9.503e+04 on 1 and 490 DF,  p-value: < 2.2e-16
```

i) Fitted model using MA method:

$$\hat{X}_t = -0.5817 + 1.0016 * m_t$$

Fitted model using loess smoothing method:

$$\hat{X}_t = 0.3878 + 0.9989 * m_t$$

ii) The trend component is significant at 95% confidence level under both method.

iii) I think the trend component is a good predictor of CO2 levels for two reasons: first, as we can see from the trend component plot in part b) that the trend is linear over the entire time range of the time series data; second, from the output of the linear models we observe that $R^2 > 0.99$f ro both models. This indicates that majority of the variation within the time series is explained by the trend component.