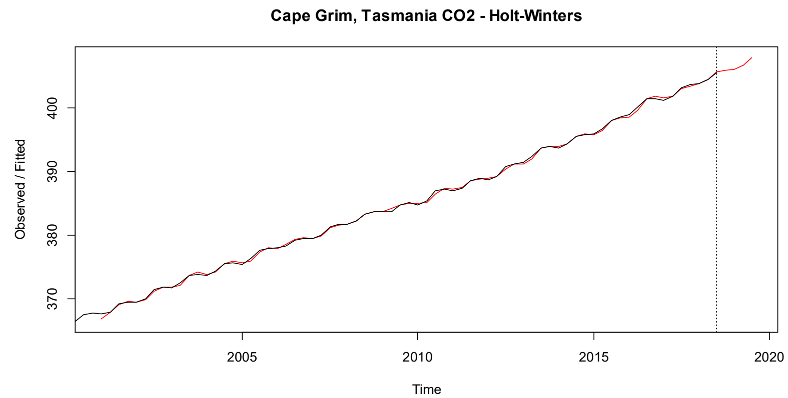
STATS 326
Applied Time Series
ASSIGNMENT TWO
ANSWER GUIDE

<u>Question One</u>:

```
> HW.fit = HoltWinters(red.CO2.ts)
> HW.fit
Holt-Winters exponential smoothing with trend and additive seasonal
   component.

Call:
HoltWinters(x = red.CO2.ts)

Smoothing parameters:
 alpha: 0.9267355
 beta : 0.0813906
 gamma: 1

Coefficients:
          [,1]
a  405.1011655
b    0.5889777
s1   0.2311236
s2  -0.2791367
s3  -0.2150001
s4   0.4588345


> HW.pred = predict(HW.fit,n.ahead=4)
> HW.pred
         Qtr1     Qtr2     Qtr3     Qtr4
2018                            405.9213
2019 406.0000 406.6531 407.9159

> actual
2018.4 2019.1 2019.2 2019.3
405.83 405.73 406.71 408.25

> HW.RMSEP = sqrt(1/4*sum((actual-HW.pred)^2))
> HW.RMSEP
[1] 0.2214015
```

```
> plot(HW.fit,HW.pred,main="Cape Grim, Tasmania CO2 - Holt-Winters")
```



The plot of the Holt-Winters model shows an excellent fit with the model (red line) being very close to the actual observations (black line). There is a slightly higher peak in 2016.

The predictions are above the actual values for 2018.4 and 2019.1 and below for 2019.2 and 2019.3.

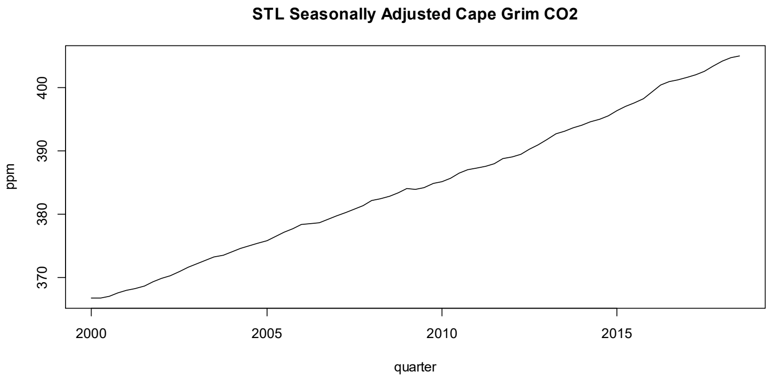
The RMSEP shows that, on average, the prediction error is 0.22 ppm.

Question Two:

```
> STL.CapeGrim$time.series[1:4,1]
[1] -0.3906880 -0.3028320  0.4787971  0.2147230
```

The seasonal peak is in the 3rd quarter with the seasonal trough in the 1st quarter.

```
> STL.red.CO2.ts = red.CO2.ts-STL.CapeGrim$time.series[,1]
> plot(STL.red.CO2.ts,main="STL Seasonally Adjusted Cape Grim
  CO2",xlab="quarter",ylab="ppm")
```



The seasonally adjusted plot is reasonably linear but the slope increases around time point 50 (2012.2).

```
> STL.fit1 = lm(STL.red.CO2.ts[-1]~red.Time[-1]+red.Time.break[-1]+
  STL.red.CO2.ts[-75])
> summary(STL.fit1)

Call:
lm(formula = STL.red.CO2.ts[-1] ~ red.Time[-1] + red.Time.break[-1] +
    STL.red.CO2.ts[-75])

Residuals:
     Min       1Q   Median       3Q      Max
-0.51687 -0.13691  0.01665  0.11813  0.51237

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         113.68138   29.12337   3.903 0.000216 ***
red.Time[-1]          0.15131    0.03826   3.955 0.000181 ***
red.Time.break[-1]    0.04304    0.01189   3.620 0.000554 ***
STL.red.CO2.ts[-75]   0.68994    0.07972   8.654 1.14e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.189 on 70 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 8.543e+04 on 3 and 70 DF,  p-value: < 2.2e-16
```

```
> t.76.stl.pred = STL.fit1$coef[1]+STL.fit1$coef[2]*76+STL.fit1$coef[3]*26+
  STL.fit1$coef[4]*STL.red.CO2.ts[75]
> t.76.stl.pred
(Intercept)
   405.7808
> t.76.pred = t.76.stl.pred+STL.CapeGrim$time.series[4,1]
> t.76.pred
(Intercept)
   405.9955

> t.77.stl.pred = STL.fit1$coef[1]+STL.fit1$coef[2]*77+STL.fit1$coef[3]*27+
  STL.fit1$coef[4]*t.76.stl.pred
> t.77.stl.pred
(Intercept)
   406.4578
> t.77.pred = t.77.stl.pred+STL.CapeGrim$time.series[1,1]
> t.77.pred
(Intercept)
   406.0671

> t.78.stl.pred = STL.fit1$coef[1]+STL.fit1$coef[2]*78+STL.fit1$coef[3]*28+
  STL.fit1$coef[4]*t.77.stl.pred
> t.78.stl.pred
(Intercept)
   407.1193
> t.78.pred = t.78.stl.pred+STL.CapeGrim$time.series[2,1]
> t.78.pred
(Intercept)
   406.8164

> t.79.stl.pred = STL.fit1$coef[1]+STL.fit1$coef[2]*79+STL.fit1$coef[3]*29+
  STL.fit1$coef[4]*t.78.stl.pred
> t.79.ma.pred
(Intercept)
   407.7709
> t.79.pred = t.79.stl.pred+STL.CapeGrim$time.series[3,1]
> t.79.pred
(Intercept)
   408.2488

> STL.pred = c(t.76.pred,t.77.pred,t.78.pred,t.79.pred)
> names(STL.pred) = c("2018.4","2019.1","2019.2","2019.3")
> STL.pred
  2018.4   2019.1   2019.2   2019.3
405.9955 406.0671 406.8164 408.2488

> STL.RMSEP = sqrt(1/4*sum((actual-STL.pred)^2))
> STL.RMSEP
[1] 0.1951761
```

The RMSEP for a seasonally adjusted MA model was 0.23 ppm

Question Three:

The seasonal estimates show that the CO2 concentration is below the overall trend for the first 2 quarters with Quarter 1 being the lowest (-0.39) and above the overall trend in the last 2 quarters with Quarter 3 being the highest (0.48).

The plot of the seasonally adjusted series shows an increasing reasonably linear trend. There is a break in the trend at time point 50 (2012.2)

The final model included a trend term, a trend break term and a lagged response to take care of autocorerelation detected in the Residual Series.

For the final model, the Residual Series appears to be reasonably random scatter about 0 with a slight positive trend for the first 2 – 3 years. There is a large negative residual for time period 38 (2009.2) and a large positive residual for time period 66 (2016.2). The plot of the autocorrelation function of the residuals shows lags 1, 11 and 16 are weakly significant, but of no real concern. The residuals appear to be normally distributed (Shapiro-Wilk *P-value* = 0.834) with isolated values at each end of the reasonably symmetric distribution due to the large residuals discussed above. The assumptions appear to be satisfied.

We have strong evidence against the hypothesis that the coefficient associated with the Time variable is 0 (*P-value* = 0.000181) and strong evidence that the coefficient associated with the Time.break variable is 0 (*P-value* = 0.000554). We have very strong evidence against the hypothesis of no autocorrelation (*P-value* ≈ 0).

The *F*-statistic provides extremely strong evidence against the hypothesis that none of the variables are related to the seasonally adjusted CO2 concentration (*P-value* ≈ 0). The Multiple R^2 is almost 1 (0.9997) indicating that nearly all the variation in the seasonally adjusted CO2 concentration is explained by the model.

The Residual Standard Error is 0.19 ppm so prediction intervals should be reasonably narrow. The model predictions can be relied on as the assumptions appear to be satisfied. The RMSEP for the predictions was 0.20 ppm which was smaller than that for the Moving Average model (0.23 ppm). Our predictions for 2018.4 and 2019.1-3 were:

Quarter 4: 406.00 ppm
Quarter 1: 406.07 ppm
Quarter 2: 406.82 ppm
Quarter 3: 408.25ppm

Question Four:

```
> STL.CapeGrim.Full = stl(full.CO2.ts,s.window="periodic")
> STL.CapeGrim.Full$time.series[1:4,1]
[1] -0.4048721 -0.3002874  0.4920871  0.2130726

> STL.full.CO2.ts = full.CO2.ts-STL.CapeGrim.Full$time.series[,1]
> full.Time.break = c(rep(0,49),full.Time[50:79]-full.Time[50])

> STL.fit = lm(STL.full.CO2.ts[-1]~full.Time[-1]+full.Time.break[-1]+
  STL.full.CO2.ts[-79])

> summary(STL.fit)

Call:
lm(formula = STL.full.CO2.ts[-1] ~ full.Time[-1] + full.Time.break[-1] +
    STL.full.CO2.ts[-79])

Residuals:
    Min      1Q  Median      3Q     Max
-0.5294 -0.1417  0.0186  0.1271  0.5095

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          114.64012   28.54054   4.017 0.000140 ***
full.Time[-1]          0.15269    0.03754   4.067 0.000118 ***
full.Time.break[-1]    0.04254    0.01109   3.837 0.000260 ***
STL.full.CO2.ts[-79]   0.68731    0.07813   8.797 4.02e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1896 on 74 degrees of freedom
Multiple R-squared:  0.9998,    Adjusted R-squared:  0.9997
F-statistic: 1.016e+05 on 3 and 74 DF,  p-value: < 2.2e-16

> t.80.stl.pred = STL.fit$coef[1]+STL.fit$coef[2]*80+STL.fit$coef[3]*30+
  STL.fit$coef[4]*STL.full.CO2.ts[79]
> t.80.stl.pred
(Intercept)
   408.3857
> t.80.pred = t.80.stl.pred+STL.CapeGrim.Full$time.series[4]
> t.80.pred
(Intercept)
   408.5988

> t.81.stl.pred = STL.fit$coef[1]+STL.fit$coef[2]*81+STL.fit$coef[3]*31+
  STL.fit$coef[4]*t.80.stl.pred
> t.81.stl.pred
(Intercept)
   409.0124
> t.81.pred = t.81.stl.pred+STL.CapeGrim.Full$time.series[1]
> t.81.pred
(Intercept)
   408.6075
```

```
> t.82.stl.pred = STL.fit$coef[1]+STL.fit$coef[2]*82+STL.fit$coef[3]*32+
  STL.fit$coef[4]*t.81.stl.pred
> t.82.stl.pred
(Intercept)
   409.6384
> t.82.pred = t.82.stl.pred+STL.CapeGrim.Full$time.series[2]
> t.82.pred
(Intercept)
   409.3381

> t.83.stl.pred = STL.fit$coef[1]+STL.fit$coef[2]*83+STL.fit$coef[3]*33+
  STL.fit$coef[4]*t.82.stl.pred
> t.83.stl.pred
(Intercept)
   410.2638
> t.83.pred = t.83.stl.pred+STL.CapeGrim.Full$time.series[3]
> t.83.pred
(Intercept)
   410.7559

> STL.Full.pred = c(t.80.pred,t.81.pred,t.82.pred,t.83.pred)
> names(STL.Full.pred) = c("2019.4","2020.1","2020.2","2020.3")
> STL.Full.pred
  2019.4   2020.1   2020.2   2020.3
408.5988 408.6075 409.3381 410.7559
```

The STL Seasonally Adjusted model using all the data to 2019.3 has almost exactly the same estimates to the original model.

The Multiple R-squared is 99.9 showing that almost all the variability is explained by the model.

The residual standard error is 0.19 ppm so prediction intervals should be narrow.

Our predictions for 2019.4 and 2020.1 – 3 are:

Quarter 4: 408.60 ppm
Quarter 1: 408.61 ppm
Quarter 2: 409.34 ppm
Quarter 3: 410.76 ppm