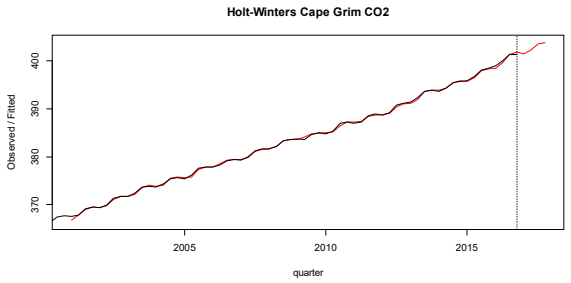
STATS 326
Applied Time Series
ASSIGNMENT TWO
ANSWER GUIDE

Question One:

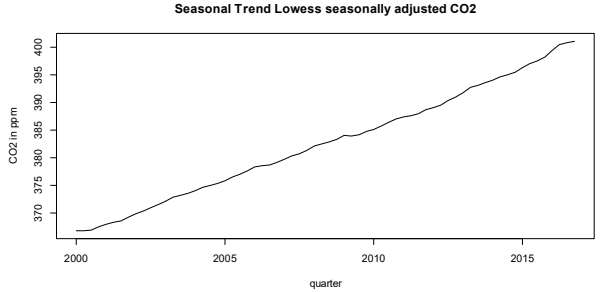
```
> HW.CapeGrim = HoltWinters(red.CO2.ts)
> HW.CapeGrim
Holt-Winters exponential smoothing with trend and additive seasonal
component.

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

Smoothing parameters:
 alpha: 0.9136498
 beta : 0.09744997
 gamma: 1

Coefficients:
           [,1]
a  401.1875008
b    0.6199149
s1  -0.2633082
s2  -0.1943729
s3   0.4707094
s4   0.2124992

> HW.CapeGrim.pred = predict(HW.CapeGrim,n.ahead=4)
> HW.CapeGrim.pred
         Qtr1     Qtr2     Qtr3     Qtr4
2017 401.5441 402.2330 403.5180 403.8797

> actual.2017 = CO2.ts[69:72]
> actual.2017
[1] 401.19 401.77 403.15 403.69

> RMSEP.HW.CapeGrim = sqrt(1/4*sum((actual.2017-HW.CapeGrim.pred)^2))
> RMSEP.HW.CapeGrim
[1] 0.3574505

> plot(HW.CapeGrim,HW.CapeGrim.pred,xlab="quarter",main="Holt-Winters Cape
  Grim CO2")
```



The Holt-Winters model fits the data well as there is little white space in the plot between the data values and the Holt Winters fit. The predictions appear consistent with the past data. The RMSEP is 0.36 ppm.

Question Two:

```
> plot(STL.CO2.ts,main="Seasonal Trend Lowess seasonally adjusted
  CO2",xlab="quarter",ylab="CO2 in ppm")
```



The plot of the seasonally adjusted data shows a reasonably linear increasing trend that appears to have a change in slope around the middle of 2011.

```
> STL.CapeGrim$time.series[1:4,1]
[1] -0.4070014 -0.3570879  0.4957070  0.2683823
```

The seasonal estimates show Quarter 1 has the lowest CO2 concentration while Quarter 3 has the highest CO2 concentration, on average.

```
> STL.CapeGrim.fit = lm(STL.CO2.ts[-1]~Time[-1]+Time.break[-1]+
  STL.CO2.ts[-68])

> summary(STL.CapeGrim.fit)

Call:
lm(formula = STL.CO2.ts[-1] ~ Time[-1] + Time.break[-1] + STL.CO2.ts[-68])

Residuals:
     Min       1Q   Median       3Q      Max
-0.45793 -0.10879 -0.01938  0.12045  0.48695

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     116.42772   30.63149   3.801 0.000328 ***
Time[-1]          0.15435    0.04024   3.836 0.000292 ***
Time.break[-1]    0.05016    0.01345   3.729 0.000414 ***
STL.CO2.ts[-68]   0.68245    0.08385   8.139 2.04e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2001 on 63 degrees of freedom
Multiple R-squared:  0.9996,    Adjusted R-squared:  0.9996
F-statistic: 5.403e+04 on 3 and 63 DF,  p-value: < 2.2e-16
```

```
> t69.stl.pred = STL.CapeGrim.fit$coef[1]+STL.CapeGrim.fit$coef[2]*69+
   STL.CapeGrim.fit$coef[3]*19+STL.CapeGrim.fit$coef[4]*STL.CO2.ts[68]
> t69.stl.pred
(Intercept)
   401.7811
> t69.pred = t69.stl.pred+STL.CapeGrim$time.series[1,1]
> t69.pred
(Intercept)
   401.3741

> t70.stl.pred = STL.CapeGrim.fit$coef[1]+STL.CapeGrim.fit$coef[2]*70+
   STL.CapeGrim.fit$coef[3]*20+STL.CapeGrim.fit$coef[4]*t69.stl.pred
> t70.stl.pred
(Intercept)
   402.4288
> t70.pred = t70.stl.pred+STL.CapeGrim$time.series[2,1]
> t70.pred
(Intercept)
   402.0717

> t71.stl.pred = STL.CapeGrim.fit$coef[1]+STL.CapeGrim.fit$coef[2]*71+
   STL.CapeGrim.fit$coef[3]*21+STL.CapeGrim.fit$coef[4]*t70.stl.pred
> t71.stl.pred
(Intercept)
   403.0753
> t71.pred = t71.stl.pred+STL.CapeGrim$time.series[3,1]
> t71.pred
(Intercept)
    403.571

> t72.stl.pred = STL.CapeGrim.fit$coef[1]+STL.CapeGrim.fit$coef[2]*72+
   STL.CapeGrim.fit$coef[3]*22+STL.CapeGrim.fit$coef[4]*t71.stl.pred
> t72.stl.pred
(Intercept)
   403.7211
> t72.pred = t72.stl.pred+STL.CapeGrim$time.series[4,1]
> t72.pred
(Intercept)
   403.9894

> STL.pred = c(t69.pred,t70.pred,t71.pred,t72.pred)
> names(STL.pred) = c("2017.1","2017.2","2017.3","2017.4")
> STL.pred
  2017.1   2017.2   2017.3   2017.4
401.3741 402.0717 403.5710 403.9894

> RMSEP.STL.CapeGrim = sqrt(1/4*sum((actual.2017-STL.pred)^2))
> RMSEP.STL.CapeGrim
[1] 0.3129915
```

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 (-0.41) and above the overall trend in the last 2 quarters with Quarter 3 being the highest (0.50).

The plot of the seasonally adjusted series shows an increasing reasonably linear trend. There is a break in the trend around the middle of 2011 with the trend becoming steeper.

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 shows a slight increasing trend at the beginning but is reasonable after observation 10. The plot of the autocorrelation function of the residuals shows lags 1 and 6 are slightly significant but as they are only just outside the confidence band there is nothing to be concerned about. The residuals appear to be normally distributed (Shapiro-Wilk *P-value* = 0.39). 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* = 0.000292) and strong evidence against the hypothesis that the coefficient associated with the Time break variable is 0 (*P-value* = 0.000414). We have very strong evidence against the hypothesis of no autocorrelation (*P-value* = 2.04×10^{-11}).

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 indicating that nearly all the variation in the seasonally adjusted CO2 concentration is explained by the model.

The Residual Standard Error is 0.2 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 2017 predictions was 0.313 which was smaller than that for the Moving Average model (0.357). Our predictions for 2017 were:

Quarter 1: 401.37
Quarter 2: 402.07
Quarter 3: 403.57
Quarter 4: 403.99

Question Four:

```
> STL.CapeGrim.F.fit = lm(STL.CO2.F.ts[-1]~Time.F[-1]+
   Time.break.F[-1]+STL.CO2.F.ts[-72])
> summary(STL.CapeGrim.F.fit)

Call:
lm(formula = STL.CO2.F.ts[-1] ~ Time.F[-1] + Time.break.F[-1] +
    STL.CO2.F.ts[-72])

Residuals:
    Min      1Q  Median      3Q     Max
-0.4514 -0.1029 -0.0292  0.1004  0.5497

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       120.25560   30.34134   3.963 0.000182 ***
Time.F[-1]          0.15975    0.03986   4.007 0.000157 ***
Time.break.F[-1]    0.04721    0.01264   3.736 0.000388 ***
STL.CO2.F.ts[-72]   0.67195    0.08306   8.090 1.65e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2006 on 67 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9996
F-statistic: 6.567e+04 on 3 and 67 DF,  p-value: < 2.2e-16


> t73.stl.pred = STL.CapeGrim.F.fit$coef[1]+STL.CapeGrim.F.fit$coef[2]*73+
   STL.CapeGrim.F.fit$coef[3]*23+
   STL.CapeGrim.F.fit$coef[4]*STL.CO2.F.ts[72]
> t73.stl.pred
(Intercept)
   404.0682
> t73.pred = t73.stl.pred+STL.CapeGrim.F$time.series[1,1]
> t73.pred
(Intercept)
   403.6555

> t74.stl.pred = STL.CapeGrim.F.fit$coef[1]+STL.CapeGrim.F.fit$coef[2]*74+
   STL.CapeGrim.F.fit$coef[3]*24+STL.CapeGrim.F.fit$coef[4]*t73.stl.pred
> t74.stl.pred
(Intercept)
   404.7232
> t74.pred = t74.stl.pred+STL.CapeGrim.F$time.series[2,1]
> t74.pred
(Intercept)
   404.3543

> t75.stl.pred = STL.CapeGrim.F.fit$coef[1]+STL.CapeGrim.F.fit$coef[2]*75+
   STL.CapeGrim.F.fit$coef[3]*25+STL.CapeGrim.F.fit$coef[4]*t74.stl.pred
> t75.stl.pred
(Intercept)
   405.3702

> t75.pred = t75.stl.pred+STL.CapeGrim.F$time.series[3,1]
> t75.pred
(Intercept)
   405.8634

> t76.stl.pred = STL.CapeGrim.F.fit$coef[1]+STL.CapeGrim.F.fit$coef[2]*76+
   STL.CapeGrim.F.fit$coef[3]*26+STL.CapeGrim.F.fit$coef[4]*t75.stl.pred
> t76.stl.pred
(Intercept)
    406.012
> t76.pred = t76.stl.pred+STL.CapeGrim.F$time.series[4,1]
> t76.pred
(Intercept)
   406.3005

> STL.F.pred = c(t73.pred,t74.pred,t75.pred,t76.pred)
> names(STL.F.pred) = c("2018.1","2018.2","2018.3","2018.4")
> STL.F.pred
  2018.1   2018.2   2018.3   2018.4
403.6555 404.3543 405.8634 406.3005
```

The model including the 2017 data has very similar estimates to our previous model (2000 – 2016). The Residual Standard Error is small (0.2 ppm) so the prediction intervals should be narrow. Our predictions should be reliable.