Question One:

```
> summary(SF.CapeGrim.fit)


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

Residuals:
     Min      1Q   Median      3Q      Max
-0.48589 -0.12979  0.01026  0.12082  0.45898

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      105.26875   30.24513   3.481 0.000939 ***
Time[-1]           0.14021    0.03972   3.530 0.000804 ***
Time.break[-1]     0.04703    0.01317   3.572 0.000707 ***
Quarter[-1]2       0.44507    0.08017   5.552 6.79e-07 ***
Quarter[-1]3       1.15836    0.07875  14.709  < 2e-16 ***
Quarter[-1]4       0.39766    0.07010   5.673 4.30e-07 ***
red.CO2.ts[-68]    0.71163    0.08275   8.600 4.68e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1925 on 60 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9996
F-statistic: 2.927e+04 on 6 and 60 DF,  p-value: < 2.2e-16

> t69.sf.pred = SF.CapeGrim.fit$coef[1]+SF.CapeGrim.fit$coef[2]*69+
  SF.CapeGrim.fit$coef[3]*19+SF.CapeGrim.fit$coef[7]*red.CO2.ts[68]
> t69.sf.pred
(Intercept)
    401.487

> t70.sf.pred = SF.CapeGrim.fit$coef[1]+SF.CapeGrim.fit$coef[2]*70+
  SF.CapeGrim.fit$coef[3]*20+SF.CapeGrim.fit$coef[4]+
  SF.CapeGrim.fit$coef[7]*t69sf.pred
> t70.sf.pred
(Intercept)
   402.1811

> t71.sf.pred = SF.CapeGrim.fit$coef[1]+SF.CapeGrim.fit$coef[2]*71+
  SF.CapeGrim.fit$coef[3]*21+SF.CapeGrim.fit$coef[5]+
  SF.CapeGrim.fit$coef[7]*t70sf.pred
> t71.sf.pred
(Intercept)
   403.5757

> t72.sf.pred = SF.CapeGrim.fit$coef[1]+SF.CapeGrim.fit$coef[2]*72+
  SF.CapeGrim.fit$coef[3]*22+SF.CapeGrim.fit$coef[6]+
  SF.CapeGrim.fit$coef[7]*t71sf.pred
> t72.sf.pred
(Intercept)
   403.9947
```

```
> SF.pred = c(t69.sf.pred,t70.sf.pred,t71.sf.pred,t72.sf.pred)
> names(SF.pred) = c("2017.1","2017.2","2017.3","2017.4")
> SF.pred
  2017.1   2017.2   2017.3   2017.4
401.4870 402.1811 403.5757 403.9947

> RMSEP.SF.CapeGrim = sqrt(1/4*sum((actual.2017-SF.pred)^2))
> RMSEP.SF.CapeGrim
[1] 0.3644395
```

The Seasonal Factor model included a Time variable, a time break variable, a seasonal factor and a lagged response variable. The Residual Series showed a slight upward trend for the first 10 observations but was reasonable after that. The plot of the autocorrelation function for the Residual Series showed significant correlations at lags 1, 11 and 16. The residuals appeared to follow a normal distribution (Shapiro-Wilk *P-value* = 0.973).

Quarters 2 – 4 CO_2 concentrations were all higher than the omitted baseline level (Quarter 1) with Quarter 3 being the highest (1.16 ppm).

The RMSEP was 0.364

Question Two:

```
> summary(Cosine.CapeGrim.fit)

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

Residuals:
    Min      1Q   Median      3Q     Max
-0.53253 -0.11442 -0.04936  0.12633  0.40090

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     125.89258   23.02993   5.466 8.68e-07 ***
Time[-1]          0.16675    0.03030   5.504 7.53e-07 ***
Time.break[-1]    0.05389    0.01136   4.743 1.28e-05 ***
Seasonal[-1]      0.56793    0.03921  14.486  < 2e-16 ***
red.CO2.ts[-68]   0.65654    0.06304  10.414 3.04e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2059 on 62 degrees of freedom
Multiple R-squared:  0.9996,    Adjusted R-squared:  0.9996
F-statistic: 3.84e+04 on 4 and 62 DF,  p-value: < 2.2e-16

> t69.cos.pred = Cosine.CapeGrim.fit$coef[1]+
  Cosine.CapeGrim.fit$coef[2]*69+Cosine.CapeGrim.fit$coef[3]*19+
  Cosine.CapeGrim.fit$coef[4]*cos((2*pi*(69-3))/4)+
  Cosine.CapeGrim.fit$coef[5]*red.CO2.ts[68]
> t69.cos.pred
(Intercept)
   401.3885

> t70.cos.pred = Cosine.CapeGrim.fit$coef[1]+
  Cosine.CapeGrim.fit$coef[2]*70+Cosine.CapeGrim.fit$coef[3]*20+
  Cosine.CapeGrim.fit$coef[4]*cos((2*pi*(70-3))/4)+
  Cosine.CapeGrim.fit$coef[5]*t69.cos.pred
> t70.cos.pred
(Intercept)
   402.1695

> t71.cos.pred = Cosine.CapeGrim.fit$coef[1]+
  Cosine.CapeGrim.fit$coef[2]*71+Cosine.CapeGrim.fit$coef[3]*21+
  Cosine.CapeGrim.fit$coef[4]*cos((2*pi*(71-3))/4)+
  Cosine.CapeGrim.fit$coef[5]*t70.cos.pred
> t71.cos.pred
(Intercept)
   403.4708

> t72.cos.pred = Cosine.CapeGrim.fit$coef[1]+
  Cosine.CapeGrim.fit$coef[2]*72+Cosine.CapeGrim.fit$coef[3]*22+
  Cosine.CapeGrim.fit$coef[4]*cos((2*pi*(72-3))/4)+
  Cosine.CapeGrim.fit$coef[5]*t71.cos.pred
> t72.cos.pred
(Intercept)
   403.9779
```

```
> Cos.pred = c(t69.cos.pred,t70.cos.pred,t71.cos.pred,t72.cos.pred)
> names(Cos.pred) = c("2017.1","2017.2","2017.3","2017.4")
> Cos.pred
  2017.1   2017.2   2017.3   2017.4
401.3885 402.1695 403.4708 403.9779

> RMSEP.Cos.CapeGrim = sqrt(1/4*sum((actual.2017-Cos.pred)^2))
> RMSEP.Cos.CapeGrim
[1] 0.3101495
```

The single Cosine model was the best predicting Harmonic model. The Residual Series showed a slight upward trend for the first 10 observations but was reasonable after that. The plot of the autocorrelation function for the Residual Series showed a barely significant correlation at lag 5 but is sufficiently small that it can be ignored. The residuals appeared to follow a normal distribution (Shapiro-Wilk *P-value* = 0.275). The RMSEP was 0.31

The Full Harmonic model produced the same results as the Seasonal Factor model, as was expected. The cosine harmonic with frequency ¼ was not significant so a Reduced Harmonic model was run without that harmonic term. The diagnostics were very similar to the Full Harmonic model. The RMSEP was slightly lower than the Full Harmonic model at 0.344

A model retaining all pairs where 1 harmonic from the pair was significant is the same as the Full Harmonic model.

Question Three:

The seasonal pattern in the CO2 concentration data is reasonably smooth. The first peak occurs at the 3rd quarter so the time component in the cosine curve was adjusted accordingly.

The Residual Series shows a slight upward trend for the first 10 observations but appears reasonable after that. The plot of the autocorrelation function for the Residual Series shows a very weak negative autocorrelation at lag 5. Since it is only just below the confidence band it can be ignored. The residuals appear to follow a normal distribution (Shapiro-Wilk *P-value* = 0.28).

We have very strong evidence that the Time variable (*P-value* = 7.53×10^{-7}) and the Time break variable (*P-value* = 1.28×10^{-5}) are not equal to 0.

We have extremely strong evidence against the hypothesis that the amplitude of the Cosine Harmonic term is 0 (*P-value* ≈ 0).

We have extremely strong evidence against the hypothesis of no autocorrelation (*P-value* = 3.04×10^{-15}).

The F-statistic provides extremely strong evidence against the hypothesis that none of the variables are related to the CO2 concentration (*P-value* ≈ 0). The Multiple R^2 is almost 1 indicating that nearly all the variation in the 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 0.31 which was smaller than the Seasonal Factor and Full Harmonic models (0.364) and the Reduced Harmonic model (0.344). Our predictions for 2017 were (in ppm):

Quarter 1: 401.39
Quarter 2: 402.17
Quarter 3: 403.47
Quarter 4: 403.98

Question Four:

```
> Seasonal.F = cos((2*pi*(Time.F-3))/4)

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

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

Residuals:
     Min       1Q   Median       3Q      Max
-0.54086 -0.11554 -0.03662  0.11992  0.45400

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     123.49375   22.30337   5.537 5.74e-07 ***
Time.F[-1]        0.16397    0.02935   5.587 4.72e-07 ***
Time.break.F[-1]  0.04867    0.01017   4.787 9.93e-06 ***
Seasonal.F[-1]    0.57480    0.03778  15.214  < 2e-16 ***
CO2.ts[-72]       0.66309    0.06106  10.860 2.47e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2025 on 66 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9996
F-statistic: 4.845e+04 on 4 and 66 DF,  p-value: < 2.2e-16

> t73.cos.pred = Cosine.CapeGrim.F.fit$coef[1]+
  Cosine.CapeGrim.F.fit$coef[2]*73+Cosine.CapeGrim.F.fit$coef[3]*23+
  Cosine.CapeGrim.F.fit$coef[4]*cos((2*pi*(73-3))/4)+
  Cosine.CapeGrim.F.fit$coef[5]*CO2.ts[72]
> t73.cos.pred
(Intercept)
   403.6894

> t74.cos.pred = Cosine.CapeGrim.F.fit$coef[1]+
  Cosine.CapeGrim.F.fit$coef[2]*74+Cosine.CapeGrim.F.fit$coef[3]*24+
  Cosine.CapeGrim.F.fit$coef[4]*cos((2*pi*(74-3))/4)+
  Cosine.CapeGrim.F.fit$coef[5]*t73.cos.pred
> t74.cos.pred
(Intercept)
   404.4765

> t75.cos.pred = Cosine.CapeGrim.F.fit$coef[1]+
  Cosine.CapeGrim.F.fit$coef[2]*75+Cosine.CapeGrim.F.fit$coef[3]*25+
  Cosine.CapeGrim.F.fit$coef[4]*cos((2*pi*(75-3))/4)+
  Cosine.CapeGrim.F.fit$coef[5]*t74.cos.pred
> t75.cos.pred
(Intercept)
   405.7858

> t76.cos.pred = Cosine.CapeGrim.F.fit$coef[1]+
  Cosine.CapeGrim.F.fit$coef[2]*76+Cosine.CapeGrim.F.fit$coef[3]*26+
  Cosine.CapeGrim.F.fit$coef[4]*cos((2*pi*(76-3))/4)+
  Cosine.CapeGrim.F.fit$coef[5]*t75.cos.pred
> t76.cos.pred
(Intercept)
   406.2918
```

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.

Question Five:

The best predicting model is the Cosine Harmonic as it has the lowest RMSEP (0.31). There is only one slightly significant lag in the acf of the residuals (lag 5) but it is only just significant so can be ignored. The assumptions are satisfied so we should be able to rely on any predictions.