Time Series: Homework 4

Paul Harmon and Jacob Dym and Justin Gomez

September 21, 2016

Report all R code either inline or in an appendix.

1) *Provide a reason for your choice of location. Report any missing observations and the range of years where you are modeling.*

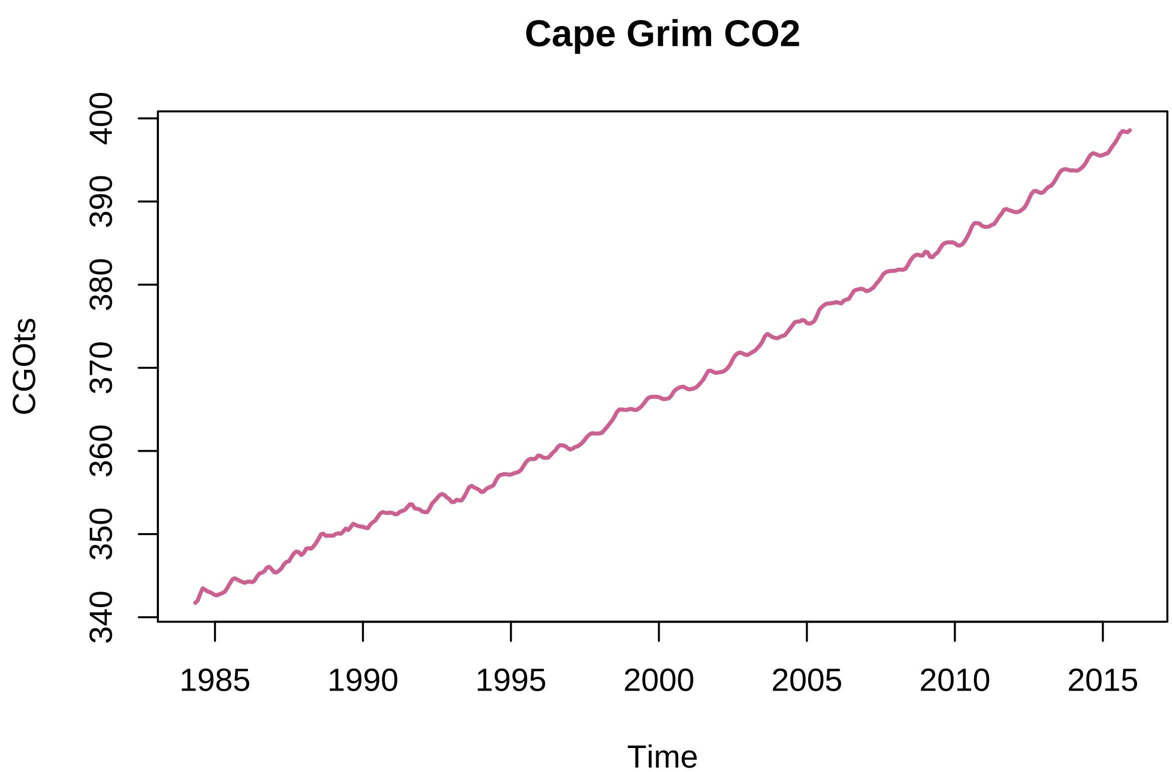
We chose a location in Australia (CGO) called Cape Grim, which is the northern-most point of Tasmania. According to the station's website (managed by Massachussetts Institute of Technology), the station there is located in the path of large air masses that settle over Antarctica and the Southern Ocean that surrounds it. Thus, measurements taken there tend to give some information about CO2 concentrations in less accessible places such as the South Pole.

The dataset contains observations over a twenty-one year period from 1984 to 2015. Thankfully, there were no missing observations in the dataset.

2) *Make a nice looking time series plot of the CO2 concentrations.*

The plot of CO2 concentrations is given below. It appears that they are increasing over time at Cape Grim.

```
plot(CGOts, col = "hotpink3", lwd = 2, main = "Cape Grim CO2")
```

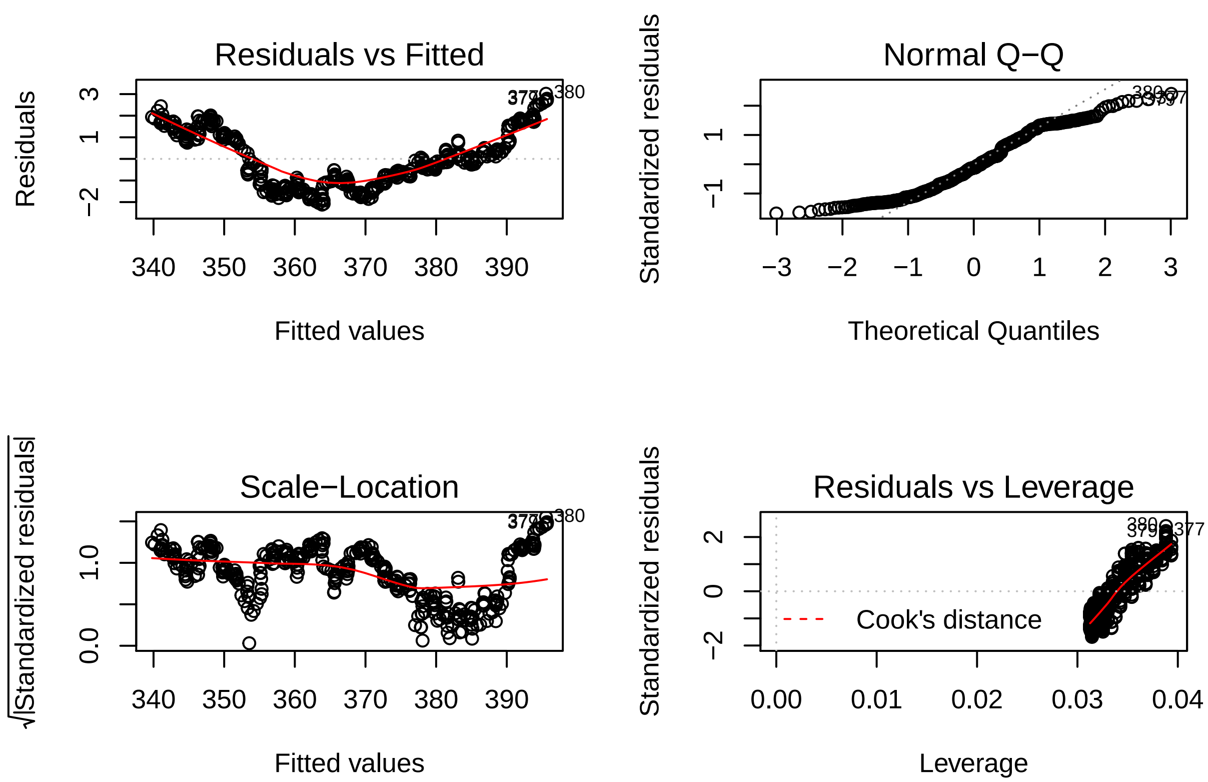


3) *Fit a linear trend plus seasonal means model to the data. Report and discuss the four panel residual diagnostics. Also make a plot of residuals vs time and discuss any potential missed pattern versus time.*

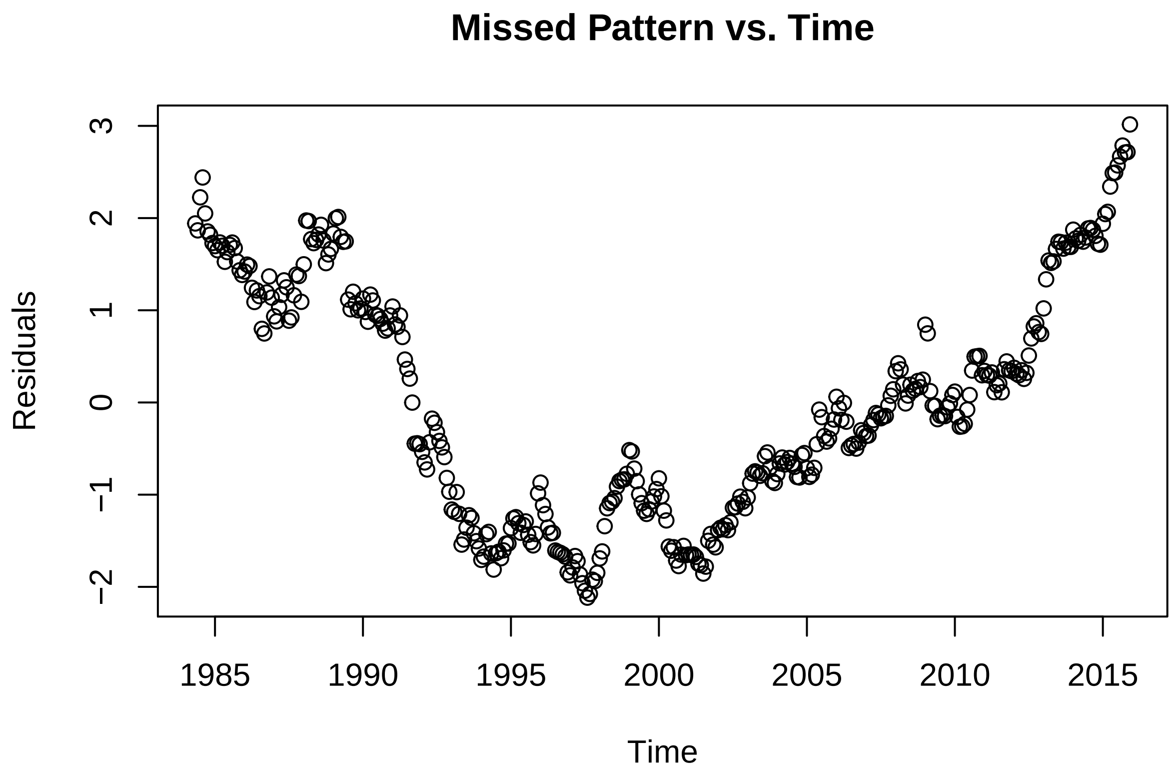
The four diagnostics are given in the plots below, as well as the plot of residuals vs. time. The Normal QQ plot examines the assumption that the resdiuals (and thus the response) are rasonably normally distributed; however this plot indicates that the tails of the distribution may be too heavy to be considered normal. The plot of residuals vs. fitted values indicates that the residuals are reasonably constantly distributed (i.e. homoskedastic) but that the response may be non-linear. This may be due to the fact that we missed some pattern over time.

Based on the plot of residuals vs. time, we can see a pattern where residuals tend to decrease and then incrase again in some sort of parabolic fashion; this pattern is missed by our model. This indicates that we ought to consider using a different model that takes this structure into account.

```
YearF <- as.vector(time(CGOts))
MonthF <- as.factor(cycle(CGOts))
CGO.lm<-lm(CGO$Value~YearF + MonthF)
par(mfrow=c(2,2))
plot(CGO.lm)
```




```
par(mfrow=c(1,1))
plot(CGO.lm$residuals~YearF,ylab="Residuals",xlab="Time", main = "Missed Pattern vs. Time")
```



4) *Provide tests for the linear and seasonal means components, conditional on each other. Report those results in two sentences including all details.*

```
LS.lm <- lm(CGO$Value~ YearF + MonthF)
anova(LS.lm)
```

```
## Analysis of Variance Table
##
## Response: CGO$Value
##            Df Sum Sq Mean Sq    F value    Pr(>F)
## YearF       1  97925   97925 59853.5218 < 2.2e-16 ***
## MonthF     11     54       5     2.9899 0.0007974 ***
## Residuals 367    600       2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
if (!require("car")) install.packages("car")
```

```
## Loading required package: car
```

```
Anova(LS.lm,type=3)
```

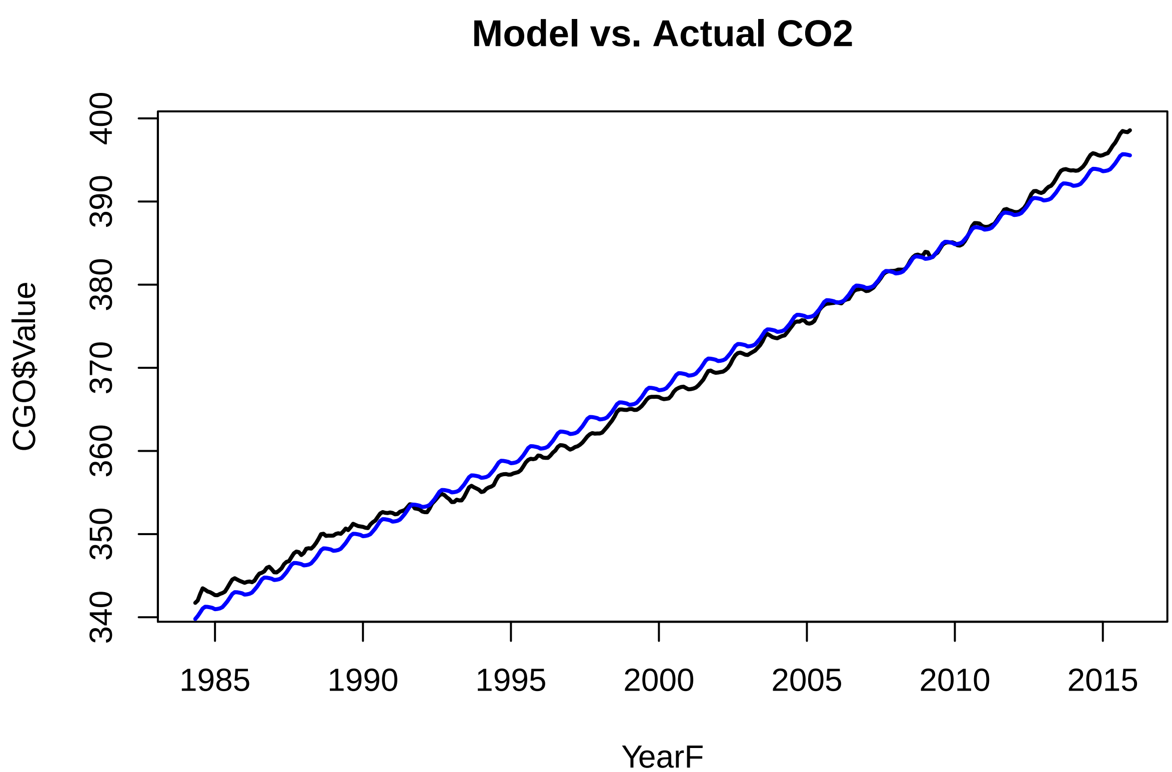
```
## Anova Table (Type III tests)
```

```
##
## Response: CGO$Value
##               Sum Sq  Df    F value    Pr(>F)
## (Intercept)  78452   1 47951.3386 < 2.2e-16 ***
## YearF        97875   1 59822.9198 < 2.2e-16 ***
## MonthF          54  11     2.9899 0.0007974 ***
## Residuals      600 367
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F test (using type III sum of squares to condition on all the covariates in the model) with 1 and 367 DF generates a F-test value of 59822.91 with a p-value of nearly 0, indicating strong evidence that the Year effect is different from 0. Further, the F test on 11 and 367 DF for the Month effect has a F-value of 2.98 yielding a p-value of 0.0008, indicating strong evidence that CO2 emissions differ by month as well.

5) *For your model, plot the original time series and the model fitted values, both versus time on the same plot. You might consider two line types or colors for the two lines. The easiest way to obtain fitted values in R is using fitted(modelname). Discuss how it appears your model does or does not describe the responses using this plot.* The plot below compares the model fitted values to the original time series. We can see our model does a great job of describing the observed CO2 values as the values are very similar.\

```r
plot(CGO$Value~YearF,type="l", main = "Model vs. Actual CO2", lwd = 2)
fits<-fitted(CGO.lm)
lines(fits~YearF,type="l", lwd = 2,col="blue")
```



6) *Document your R version.*

```r
getRversion()
```

```
## [1] '3.3.1'
```

```r
sessionInfo()$R.version$nickname
```

```
## [1] "Bug in Your Hair"
```