Solution to Series 2

1. a)
```
> hstart <-read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/hstart.dat")
> hstart <- ts(hstart[,1], start=1966, frequency=12)
```

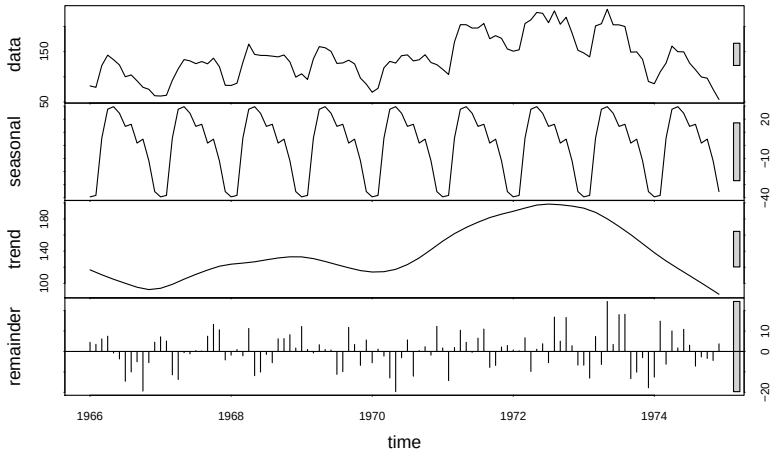
The time series under investigation is non-stationary (see the plot). Its non-stationary properties consist of a trend and deterministic seasonal fluctuations (by month).
We can decompose this time series into its trend, its seasonal fluctuations (month effects) and the remainder:

$$X_t = m_t + \alpha_{i\langle t\rangle} + E_t$$

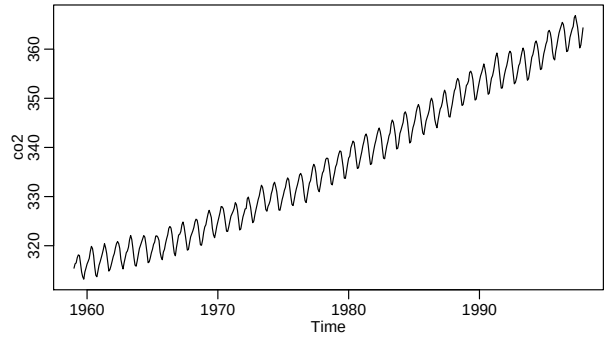
If the trend can be described by a polynomial of degree k, we then have

$$m_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \ldots + \beta_k \cdot t^k$$

b) The non-parametric STL decomposition:
```
> H.stl <- stl(hstart, s.window="periodic")
> plot(H.stl)
```



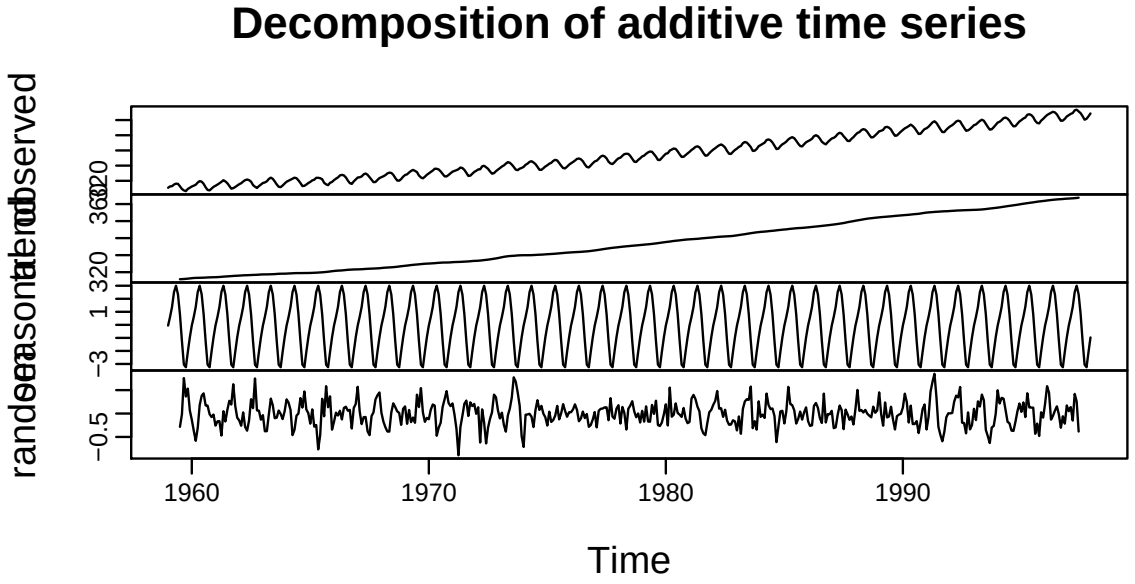
2. a)
```
> data(co2)
> plot(co2)
```



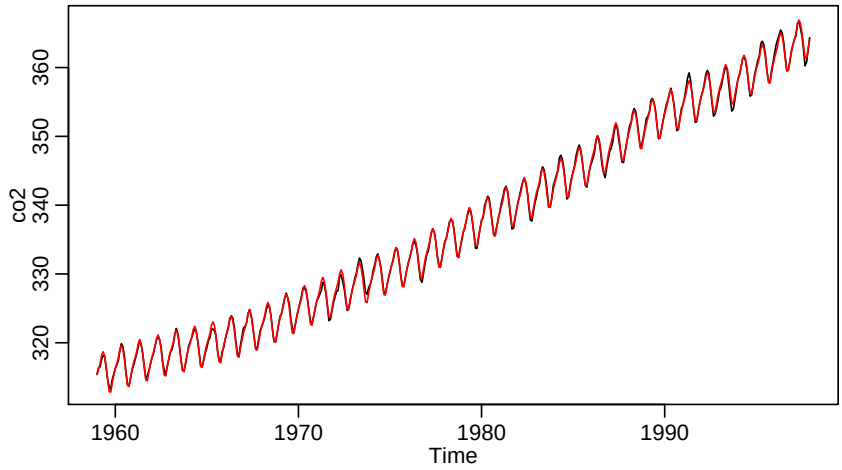
The time series under investigation is non-stationary. Its non-stationary properties consist of a trend and deterministic seasonal fluctuations (by month). We can decompose this time series into its trend, its seasonal fluctuations (month effects) and the remainder:

$$X_t = m_t + \alpha_{i\langle t\rangle} + E_t$$

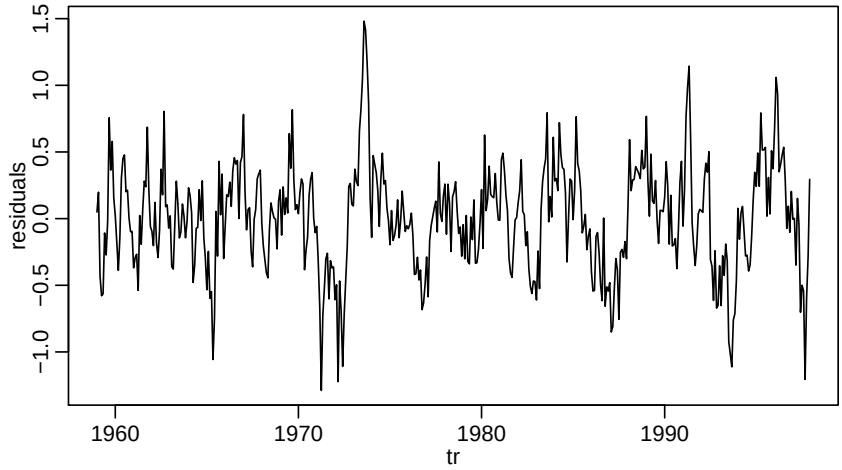
b)
```
> ts_decomp<-decompose(co2,type="additive")
> plot(ts_decomp)
```



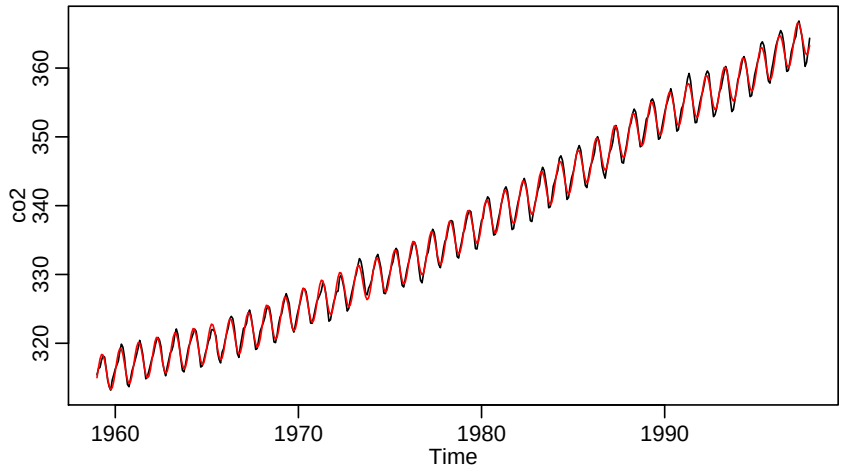
c)
```
> library(mgcv)
> tr <- as.numeric(time(co2))
> mm <- rep(c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
              "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))
> mm <- factor(rep(mm,40),levels=mm)[1:length(co2)]
> fit <- gam(co2 ~ s(tr) + mm)
> plot(co2, ylab="co2",type="l")
> lines(tr, fitted(fit), col="red")
```



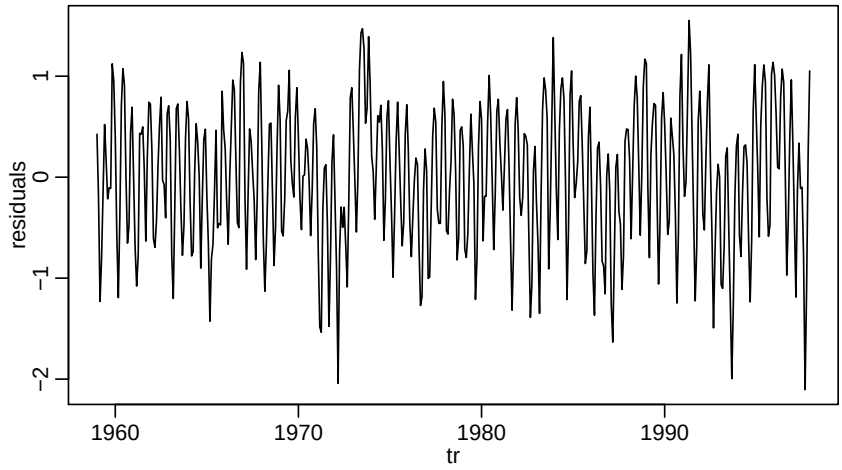

```
> plot(tr,co2-fitted(fit),ylab="residuals",type="l")
```



d)
```
> fit <- gam(co2 ~ s(tr) + sin(2 *pi*tr)+cos(2*pi*tr))
> plot(co2, ylab="co2")
> lines(tr, fitted(fit), col="red")
```




```
> plot(tr,co2-fitted(fit), ylab="residuals",type="l")
```



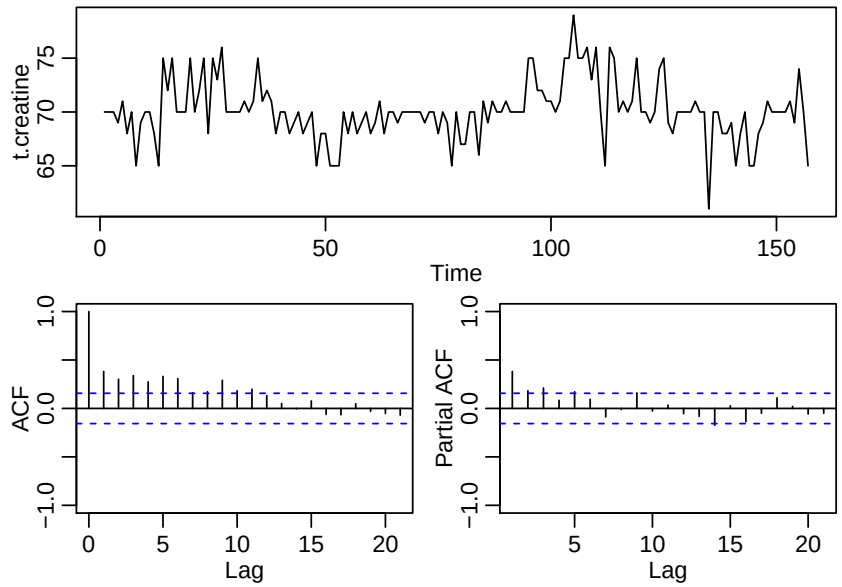
The variance of the residuals seems higher when we use the oscilator. On the other hand this time the residuals contain less structure.

3. **a)** If the analyzing machine works correctly, the measurement it takes at time t is independent of the previous measurements:

$$X_t = \mu + E_t, \qquad E_t \text{ i.i.d.},$$

where μ denotes the creatine content of a standard sample.

b)
```
> d.creatine <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/kreatin.dat",
    header = TRUE)
> t.creatine <- ts(d.creatine[, 2], start = 1, frequency = 1)
> layout(matrix(c(1, 1, 2, 3), 2, 2, byrow = TRUE))
> plot(t.creatine)
> acf(t.creatine, plot = TRUE, ylim=c(-1,1))
> pacf(t.creatine, plot = TRUE, ylim=c(-1,1))
```



Since the data exhibit strong time-based correlation, the ideal model does *not* fit.
Thus the next steps would be to check the machine (perhaps some residue from the samples remains inside) and to ensure the standard samples really are manufactured in a way that rules out correlation.
Note: If you consider this time series to be non-stationary (not having a constant level, exhibiting standard autocorrelations that decay very slowly), then you are not the only one to think this way. If this is true, the machine must in any case be regarded as useless: since the standard samples all have the same creatin content, the measurements should have constant expectation.