Assignment 4

Azoacha Forcheh, 20558994

1. **Antarctic sea ice** On the website for this assignment, in the directory called `R` you will find a file called `seaice.csv`. Download this file.

   We will need to do some manipulations to this data before it can be easily used in our analysis.

```
seaice <- read.csv("./R/seaice.csv", header=TRUE)
```

   The last line above shows that `seaice` is a data frame having four variates. The first three identify the year, month, and day that the last measurement was taken. The last measure is a determination of the extent of Antarctic sea ice in millions of square kilometres as determined by satellite imagery.

   (a) *Irregular time series* First, we begin by putting the year, month, and day together into a single date. Do this as follows:

```
date <- paste(seaice$Year,seaice$Month,seaice$Day, sep=" ")
head(date)
```

```
## [1] "1979 1 2"  "1979 1 4"  "1979 1 6"  "1979 1 8"  "1979 1 10" "1979 1 12"
```

```
# The following function turns that into a standard date format
# You will need the package "lubridate"
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```
date <- ymd(date)
head(date)
```

```
## [1] "1979-01-02" "1979-01-04" "1979-01-06" "1979-01-08" "1979-01-10"
## [6] "1979-01-12"
```

```
# And we create a new data structure called sea.ice
sea.ice <- data.frame(date=date, extent = seaice$Extent)
head(sea.ice)
```

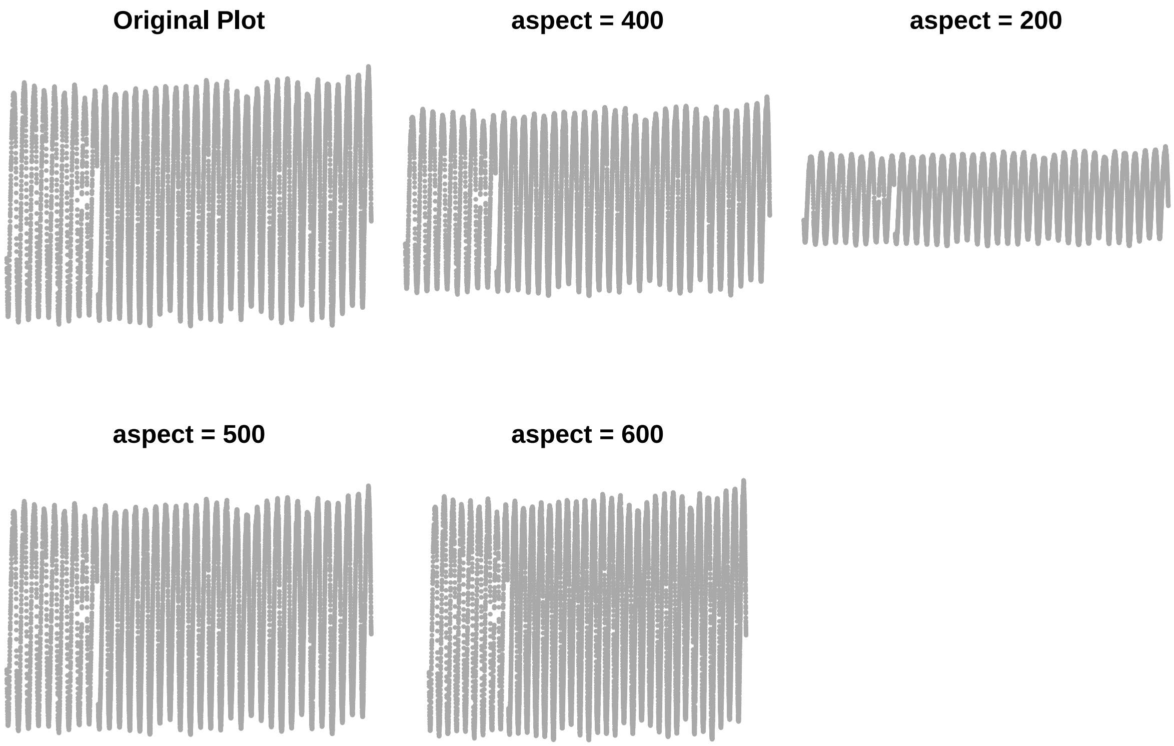
```
##         date extent
## 1 1979-01-02  6.945
## 2 1979-01-04  6.841
## 3 1979-01-06  6.648
## 4 1979-01-08  6.270
## 5 1979-01-10  6.138
## 6 1979-01-12  5.957
```

   We will be using the data frame `sea.ice` for this part of the question.

   i. **(4 marks)** Call `plot(...)` directly on the data frame `sea.ice` with `cex=0.3, pch=19, col="darkgrey"`. Zoom in on this plot and experiment with different aspect ratios by reshaping the plot's window. Describe whatever interesting characteristics you see in the time series.

```
# Aspect ratio
savePar = par(mfrow=c(2,3), mar=c(2.5,0.1,3,0.1))
plot(sea.ice, xlab="", ylab="", axes=FALSE,
     main="Original Plot", cex=0.3, pch=19, col="darkgrey")
for (aspect in c(400, 200, 500, 600)) {
```

```
  plot(sea.ice, asp=aspect,
       main=paste("aspect =",aspect),
       xlab="", ylab="", axes=FALSE,
       cex=0.3, pch=19, col="darkgrey")
}
par(savePar)
```




```
# Zooming with loon
require(loon)
```

```
## Loading required package: loon
```

```
## Loading required package: tcltk
```

```
l_plot(sea.ice, showScales=TRUE,
     size=0.3, glyph="ccircle",
     color="darkgrey")
```

```
## [1] ".l0.plot"
## attr(,"class")
## [1] "loon"
```

The value of the extent of the Antartic sea ice stays within the same range, but it continuously sharply decreasing and increasing over short time periods.

ii. Decomposition into a long term trend and a short term trend. First we need to transform the data again.

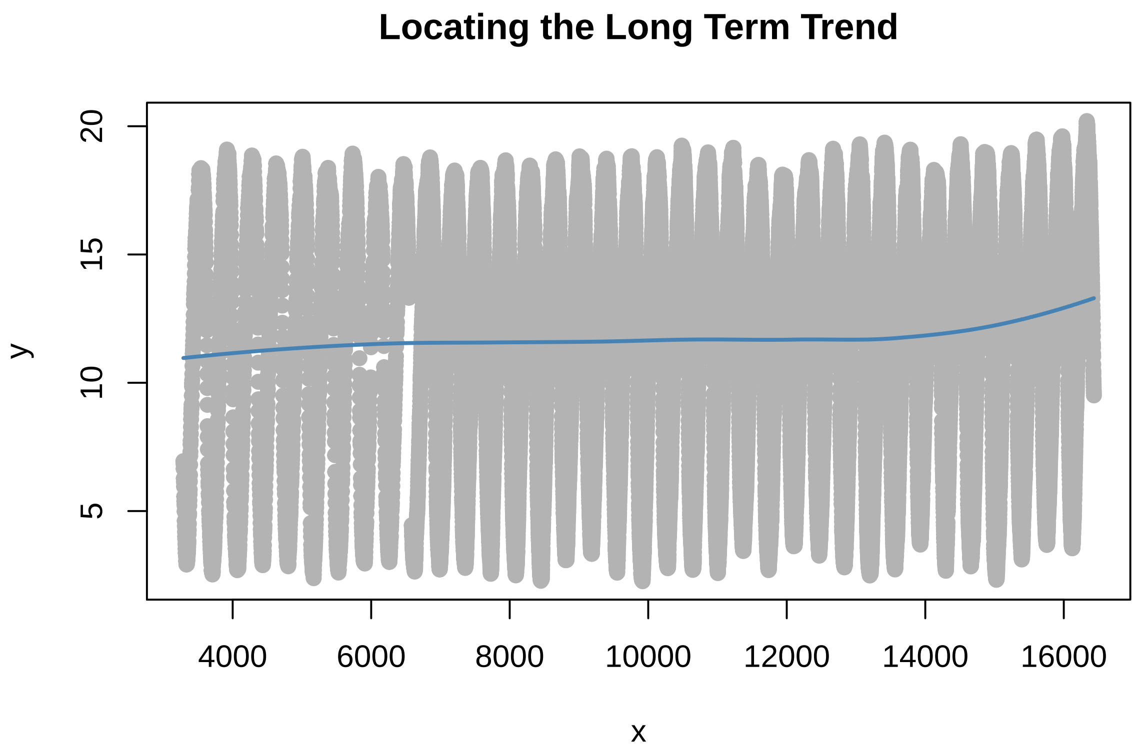
```
sea.ice.xy <- xy.coords(sea.ice)
y <- sea.ice.xy$y
x <- sea.ice.xy$x
# plotting x and y as before will yield the same plot, but
# with the dates lost.
#plot(x, y, cex=0.3, pch=19, col="darkgrey")
```

The data range from the beginning of 1979 to the end of 2014. The number of years covered then is

(2014-1979 + 1)=36. So each year is approximately 1/36 of the series. A month is about 1/12 of this again.

- **(2 marks)** Using `loess(..)` as in class, find a long term trend using `span=2/3`. Add this trend as a curve in a different colour on top the points (as in the above plot). Describe the fit.

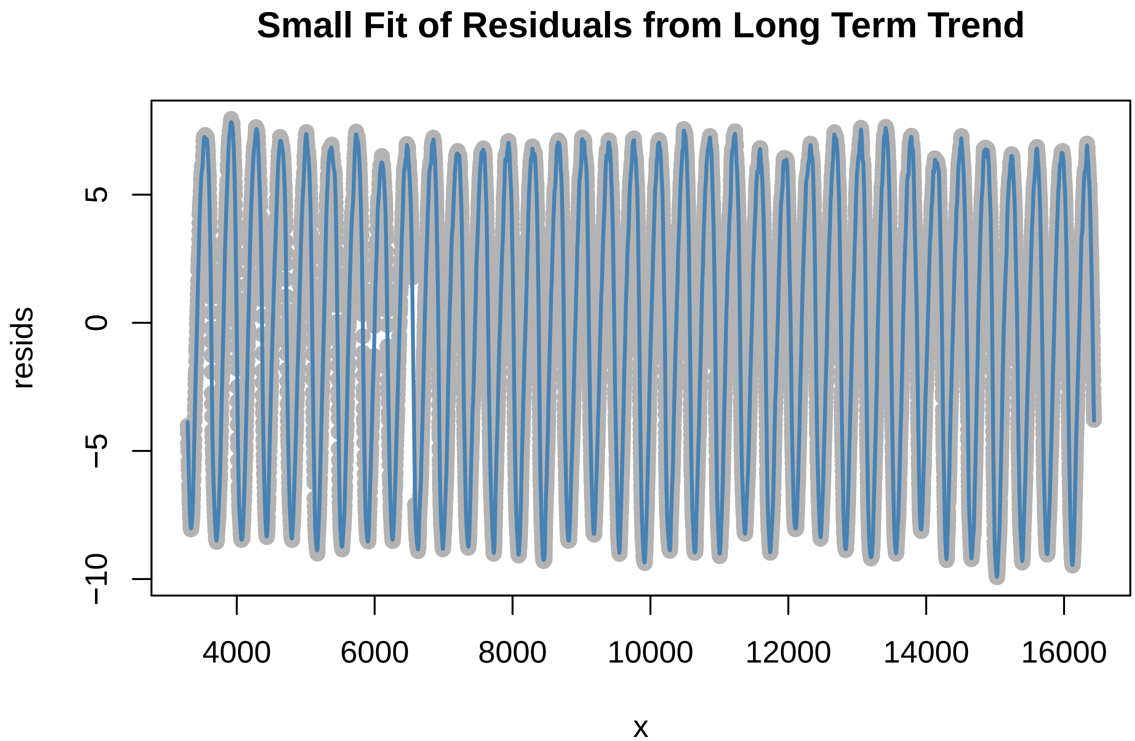
```
ltfit = loess(y~x, span=2/3)
ltpred = predict(ltfit)
xord = order(x)
plot(x, y, col="grey70", pch=19,
     main="Locating the Long Term Trend")
lines(x[xord], ltpred[xord], lwd=2, col="steelblue")
```



The extent of Antarctic sea ice stays fairly consistent and stationary, then there is a small jump upwards and the extent begins to trend upwards (i.e. increase) slowly.

- **(2 marks)** Now fit a `loess` short term smooth with `span=1/(36*12)` to the residuals from the long term trend. Plot the residuals from the long-term trend and add the fitted short term trend. Describe the fit.

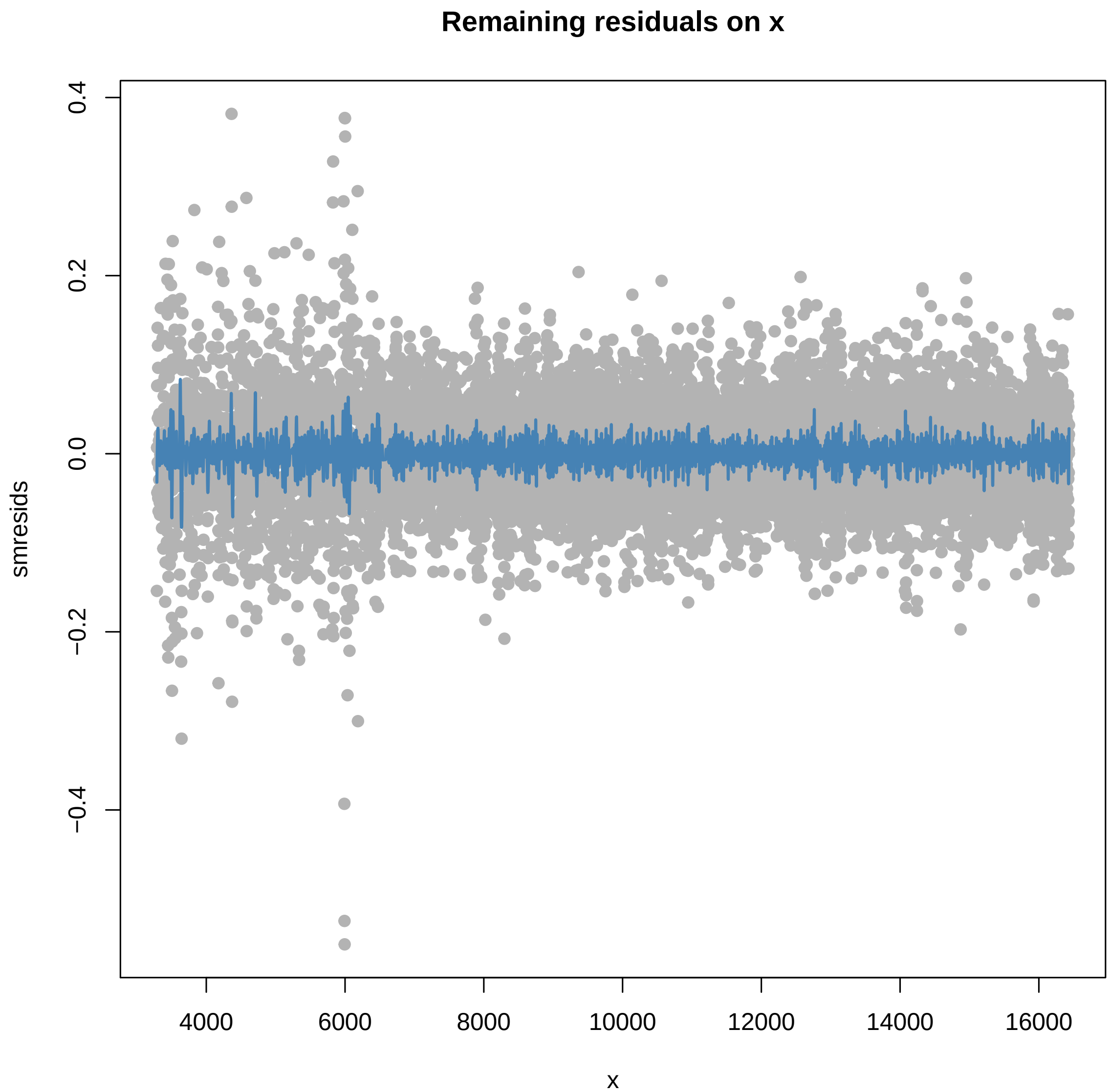
```
resids = ltfit$residuals
plot(x, resids, col="grey70", pch=19,
     main="Small Fit of Residuals from Long Term Trend")
smfit = loess(resids~x, span=1/(36*12))
smpred = predict(smfit)
lines(x[xord], smpred[xord], lwd=2, col="steelblue")
```



The residuals exhibit a consistently seasonal trend - it trends upward and the downwards very quickly for about every $333\frac{1}{3}$ units of x, staying with the range of about $[-10, 10]$.

- **(3 marks)** Plot the residuals from the short-term trend. Describe any patterns you find.

```
smresids = smfit$residuals
plot(x, smresids, col="grey70", pch=19,
     main="Remaining residuals on x")
rfit = loess(smresids~x, span=1/(36*12))
rpred = predict(rfit)
lines(x[xord], rpred[xord], lwd=2, col="steelblue")
```



There is no consistent trend in the plot - the remaining residuals are too random to locate such a distinct pattern.

- **(2 marks)** Explain your findings.

  The graphs represent the different parts of the decomposition of the time series model of the Antartic sea ice extent, with the first long term trend of the extent being the main trend in the model, the short term trend of the residuals representing the seasonality of the model, and the remaining residuals representing the remaining randomness in the model.

(b) **Monthly averages** The original series was very irregular. Here we replace it with a complete series using the average extent for each month. This time series is created as follows.

```
months <- unique(seaice$Month)
nmonths <- length(months)
years <- unique(seaice$Year)
nyears <- length(years)
# check the beginning and the end months.
seaice[1,]
```

```
##   Year Month Day Extent
## 1 1979     1   2  6.945
```

```
seaice[nrow(seaice),]
```

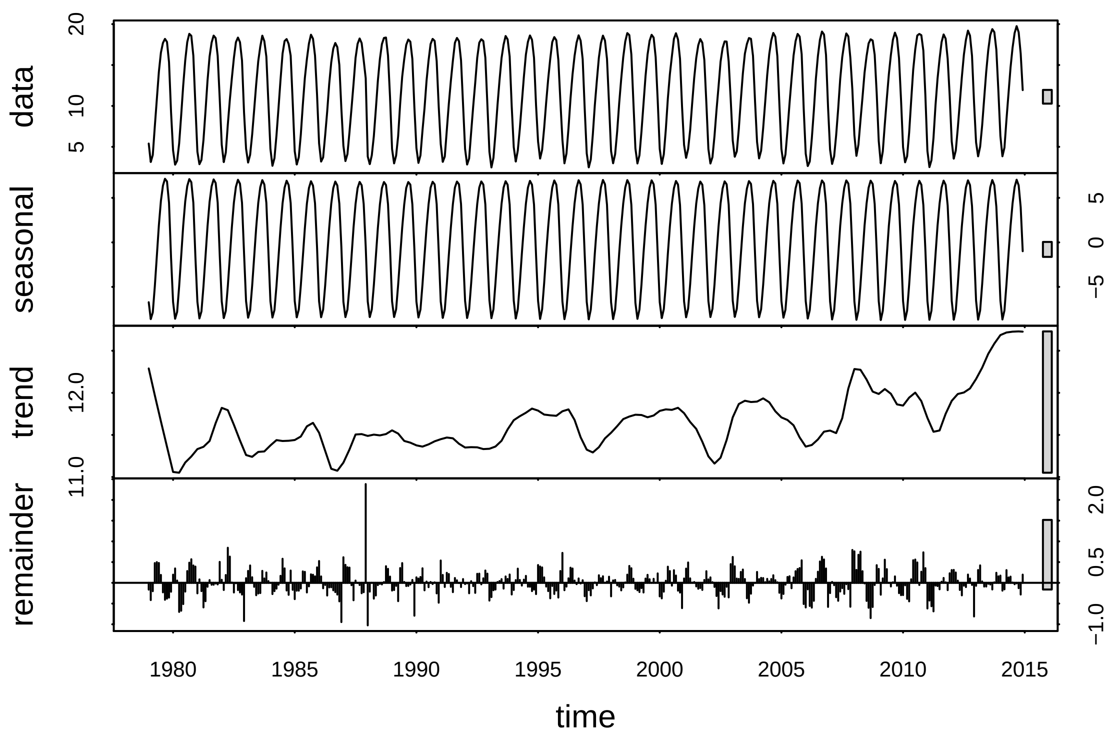
```
##       Year Month Day Extent
## 11552 2014    12  31  9.508
```

```
maxN <- nmonths * nyears
# place holder
iceMonthly <- vector("numeric", length=maxN)
for (i  in 1:nyears) {
  year <- seaice[,"Year"] == years[i]
  iceyear <- seaice[year,c("Month","Extent")]
  for (j in 1:nmonths) {
    index <- (i-1) * nmonths + j
    selection <-  iceyear[,"Month"] == months[j]
    iceMonthly[index] <- mean(iceyear[selection,"Extent"])
  }
}
#  Now we create a time series object
ice <- ts(iceMonthly, start=c(1979,1), frequency=12)
```

Conduct a seasonal trend analysis on `ice` using the built-in `stl(...)` function with parameters `s.window=11`, `s.degree=1`, and `t.degree=1`.

i. **(3 marks)** Plot the output of `stl` and describe your findings.

```
plot(stl(ice, s.window=11,
         s.degree=1, t.degree=1))
```



The long term trend of the data is cyclical - the short-term trends/cycles where the extent decreases and the increases again don't repeat regularly.

ii. **(2 marks)** Compare the output above with the decomposition of the irregular time series in part (a).

The shape of the seasonal portion of the diagram and randomness of the remainder is consistent with what was plotted in part (a) for the short term smooth of residuals and the plot of the remaining residuals respectively.

However, the main trend of the two questions are different, with the trend in part (a) having exhibited a stationary trend and the trend in part (b) being more cyclical. This is due to the irregularity of the data used in part (a), with part (b) showing the true trend of the data as the ice extents for the missing dates in part (a) have been accounted for.