

Solution to Series 1

1. a) Read in the data and define it correctly as a time series of class `ts`.

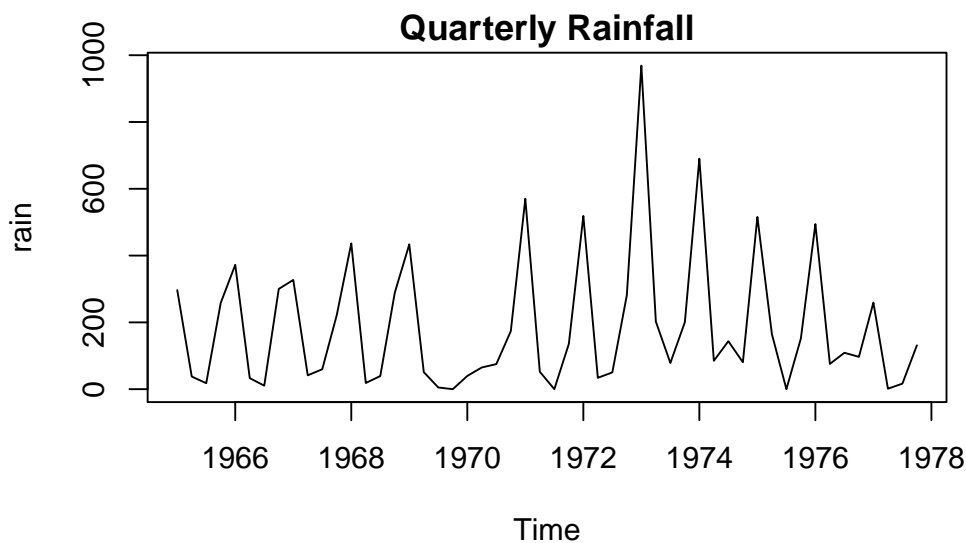
R-Hint: `ts.d <- ts(data= ..., ...)`

```
> dd <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/rain.txt", header=T)
> ts.dd <- ts(dd, start=1965, frequency=4)
> ts.dd
```

	Qtr1	Qtr2	Qtr3	Qtr4
1965	296.80	37.80	17.90	258.30
1966	372.30	32.90	10.45	300.20
1967	327.10	41.40	59.60	221.90
1968	436.50	18.50	39.20	289.20
1969	433.60	51.00	5.00	0.00
1970	39.60	65.00	75.00	173.70
1971	570.40	52.10	0.00	135.40
1972	518.70	34.00	50.20	281.30
1973	968.80	202.20	78.40	200.60
1974	690.20	85.00	143.60	80.50
1975	515.70	165.00	0.00	151.80
1976	494.30	75.60	108.95	96.80
1977	259.20	1.40	16.50	131.40

- b) Plot the time series.

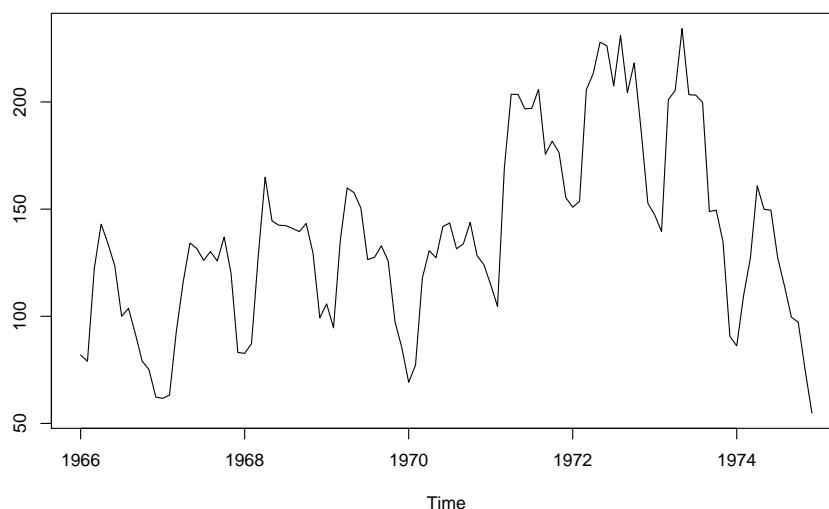
```
> plot(ts.dd, ylab="rain", main="Quarterly Rainfall")
```



2. a) Sunshine duration per month in Basel from 1990 to 2000.
This is monthly data, so the frequency is 12 and $deltat$ is $\frac{1}{12}$.
- b) Number of newborn babies in the city of Zurich per year from 2000 to 2011.
The frequency is 1, $deltat$ is 1 as well.
- c) Time Series with the number of reservations in a restaurant for every night during 4 weeks.
An obvious time unit would be one week, so frequency is 7 and $deltat$ is $\frac{1}{7}$.
- d) Time Series with the water runoff of a river. The data has been collected every day for four years.
The time unit here would be one year. Frequency is 365 and $deltat$ is $\frac{1}{365}$.

- e) Time Series with the number of reservations in a restaurant for every night during 4 years. It depends on whether we have a seasonal effect. If there are no obvious differences throughout the year (Christmas, Summer Holidays), then we can say one week is one time unit. Otherwise it would be one year. The crucial question is: "How long does it take until I get similar data?"
3. a) The series is non-stationary. There is a nonlinear trend, mostly increasing. There is no seasonal component. Log-transformation might be beneficial (see part b)).
- b) The series seems to be stationary; the trend from before was removed by differencing. We achieved an approximately constant variance by taking the logarithm. However it is known, that the Log>Returns of financial time series are NOT stationary, since there are volatility clusters, e.g. $Var(X_t) \neq Var(X_t|X_{t-1}, X_{t-2}, \dots)$, where X_t are the Log>Returns.
- c) The series is non-stationary. There is an (exponentially) increasing trend and a (multiplicative) seasonal component with period 1 year. Thus, we need to log-transform the data in order to get the usual additive components.
- d) Dito, as in Task c).
- e) The series is non-stationary. There is no trend, but a seasonal effect with period 1 year.
- f) The series is probably stationary. There are no trend and seasonal effects, but a (non-seasonal!) periodicity with a period of approximately 11 years. Maybe, the variance could be varying over time. The data show some right-skewness and hence, a log-transformation is required.
- g) The series is probably stationary. There is no clear trend (maybe a slight one), and no seasonal effect visible.
4. a)

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



The time series under investigation is non-stationary (see the plot for Part c). 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:

$$X_t = m_t + \alpha_{i(t)} + E_t$$

- b) First we remove the seasonal fluctuations by computing $Y_t = X_t - X_{t-12}$. In the next plot, we then see that a trend nonetheless remains. Taking first-order differences of Y_t , i.e.

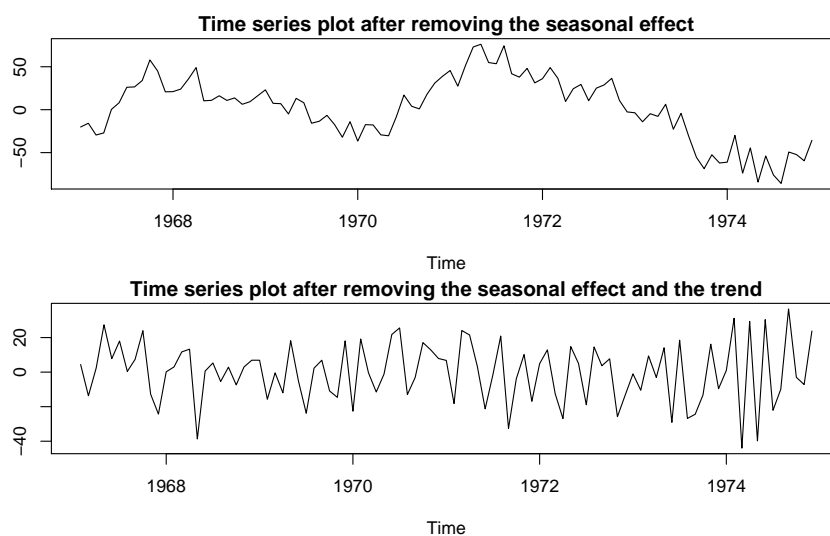
$$Z_t = Y_t - Y_{t-1} = (X_t - X_{t-12}) - (X_{t-1} - X_{t-13}),$$

finally gives us a stationary series.

```

> H.y <- diff(hstart,lag=12)
> H.z <- diff(H.y,lag=1)
> par(mfrow=c(2,1))
> plot(H.y, main="Time series plot after removing the seasonal effect")
> plot(H.z,
      main="Time series plot after removing the seasonal effect and the trend")

```

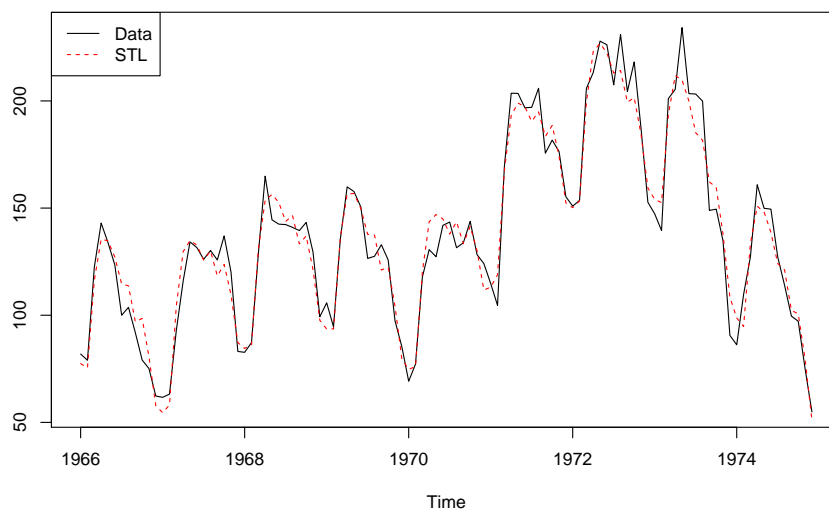


c) The non-parametric model STL fits the data better than the parametric one does.

```

> H.stl <- stl(hstart, s.window="periodic")
> remainder <- H.stl$time.series[,3]
> plot(hstart)
> H.np <- hstart - remainder
> lines(H.np, col=2, lty=2)
> legend("topleft", legend=c("Data", "STL"), col=c(1,2), lty=c(1,2))

```



In the residual time series plot is shown below. No systematic behavior is visible and hence it could be concluded that the combination of the estimated trend and the seasonal effect fits the data reasonably well.

```

> plot(ts(remainder, start=1966, freq=12), lty=3, col=2,
      ylab="Residuals", main="Residual plot")
> abline(h=0)

```

