Solution to Series 1

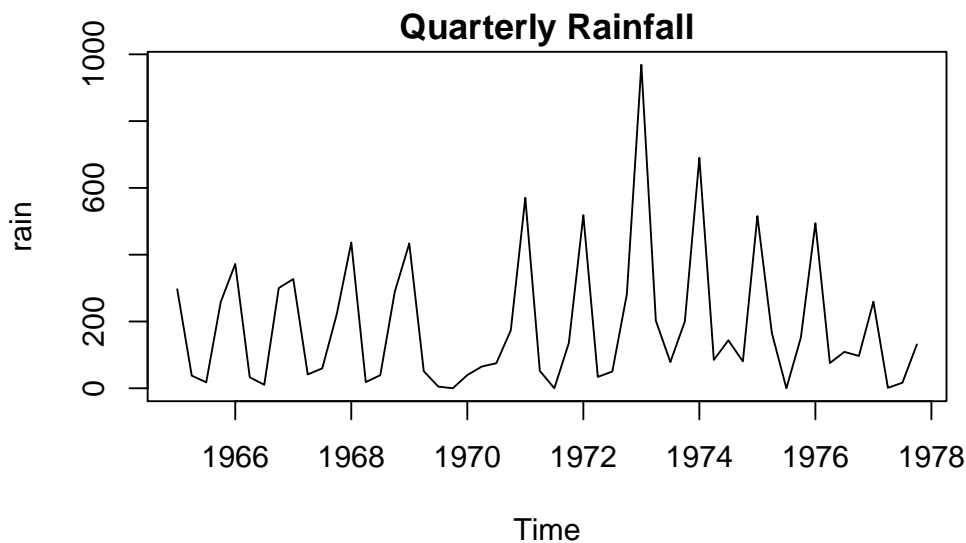**1. a)** Read in the data and define it correctly as a time series of class `ts`.

```
> 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) Read in the data `rainDay.txt` and tell R that the column `DATE` is a date.

```
> dd <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/rainDay.txt", header=T)
> dd$DATE <- as.Date(dd$DATE, format="%d.%m.%Y")
```

b) Define your data (without the `DATE`-column) correctly as a time series of class `ts`.

Since it is daily data, the frequency is 365.

```
> ts.dd <- ts(dd[,2], start=2000, freq=365)
> str(ts.dd)

 Time-Series [1:2922] from 2000 to 2008: 0 12.9 0 0.05 3.55 2.05 3.5 7.65 1.1 9.8 ...
```

c) Use the R-Functions `weekdays()`, `months()` and `quarters()` to create these factors. Combine them together with the rainfall data and the date into one dataframe.

```
> t.weekday <- factor(weekdays(dd$DATE, abbreviate=TRUE),
                      levels=c("Mon","Tue","Wed","Thu","Fri","Sat","Sun"))
> t.month <- factor(months(dd$DATE, abbreviate=TRUE), levels=
                c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep",
                  "Oct","Nov","Dec"))
> t.quarter <- quarters(dd$DATE, abbreviate=FALSE)
> dd.new  <- data.frame(date=dd$DATE, rain=ts.dd, weekday=t.weekday, month=t.month,
                quarter=t.quarter)
```

d) Plot the boxplots of the rainfall data with `weekday`, `month` and `quarter` as grouping variables. Use the `logst`-transformation (from the package `regr0`) on the `rain` variable. This function automatically adds a small constant to zero values in order to apply the log transformation.

```
> require(regr0)
> par(mfrow=c(2,2))
> boxplot(logst(dd$rain)~dd.new$month)
> boxplot(logst(dd$rain)~dd.new$weekday)
> boxplot(logst(dd$rain)~dd.new$quarter)
```
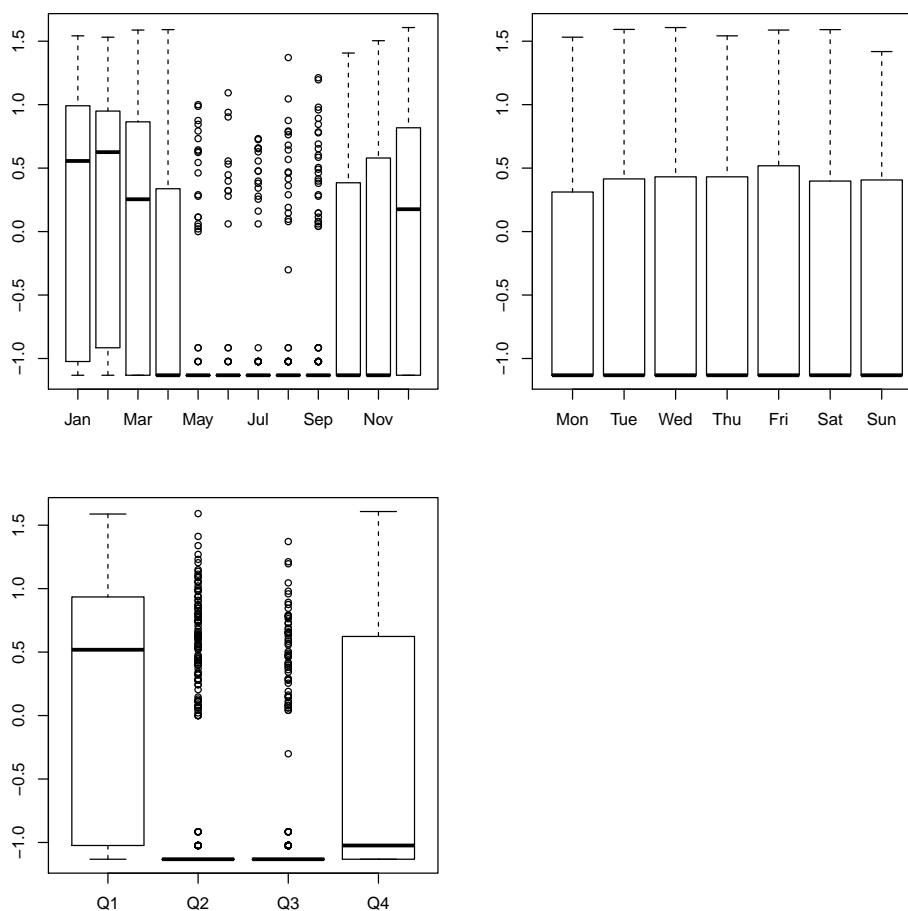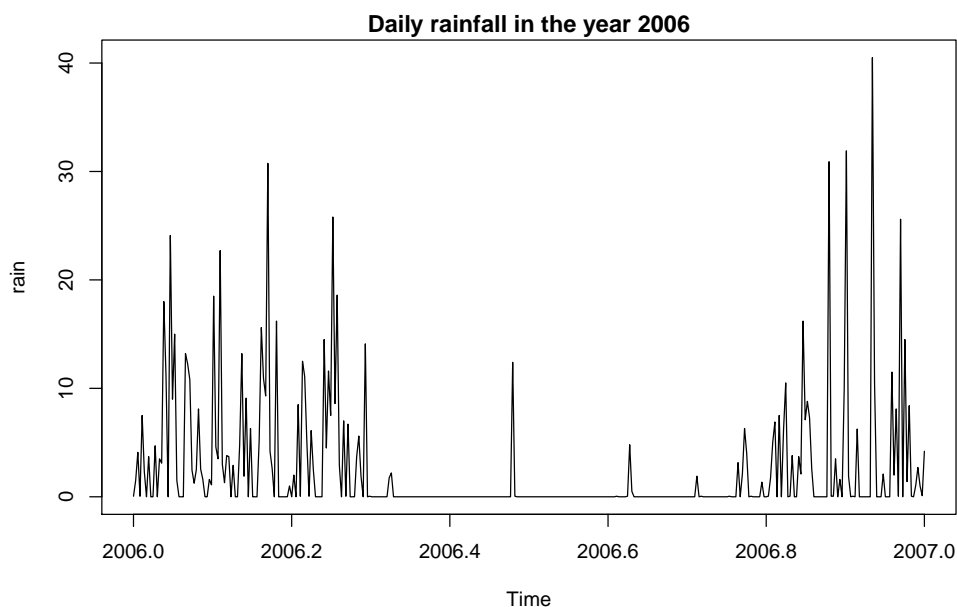

We can see from the plots that there is very little rain from May to September (Quartal 2 and 3) compared to the other months.

e) Plot the part of the time series from 2006 to 2007.

```
> plot(window(ts.dd, start=2006, end=2007), ylab="rain",
     main="Daily rainfall in the year 2006")
```

Daily rainfall in the year 2006

3. a) Sunshine duration per month in Basel from 1990 to 2000.
This is monthly data, so the frequency is 12 and `deltat` $\frac{1}{12}$.

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) 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) 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) Number of reservations in a restaurant for every night during 4 years.
It depends on wether 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?"

4. a) The series is non-stationary. There is a nonlinear trend, mostly decreasing. There is no seasonal component.

b) The trend from before was removed by differencing. We achieved an approximately constant variance by taking the logarithm. Although the series thus might seem to be stationary, it is known, that the Log-Returns of financial time series are NOT stationary, since there are volatility clusters, e.g. $Var(X_t) \neq Var(X_t|X_{t-1}, X_{t-2}, ...)$, 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.

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.

g) The series is probably stationary. There is no clear trend (maybe a slight one), and no seasonal effect visible.