Read the data

```
trainB <- read.csv("../../data/train4B.csv")
## WNV just appear in three species
table(trainB$Species, trainB$WnvPresent)
```

```
##
##                             0    1
##   CULEX ERRATICUS           1    0
##   CULEX PIPIENS          1766  184
##   CULEX PIPIENS/RESTUANS 3601  225
##   CULEX RESTUANS         2345   48
##   CULEX SALINARIUS         83    0
##   CULEX TARSALIS            6    0
##   CULEX TERRITANS         216    0
```

```
## ## We just keep these three species
## trainB <- trainB[trainB$Species %in% c("CULEX PIPIENS", "CULEX PIPIENS/RESTUANS", "CULEX RESTUANS"),
## ## WNV doesn't appear in some Addresses (39/138 addresses don't have wnv)
## address.wnv <- as.data.frame.matrix(table(trainB$Address, trainB$WnvPresent))
## wnv.address <- rownames(address.wnv[address.wnv$"1"!=0,])
## trainB <- trainB[trainB$Address %in% wnv.address, ]
## Transform the data B
trainB$Date <- as.Date(trainB$Date)
trainB$Month <- factor(trainB$Month,
                       levels=c("May", "June", "July", "August", "September", "October"))
trainB$Weekday <- factor(trainB$Weekday,
                         levels=c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"))
trainB$Block <- factor(trainB$Block)
```

NumMosquitos vs. Date grouped by Block

```
## define the plotting function
library(ggplot2)
library(plyr)
plot.ts.group <- function(data, segments, group, year) {
    ## Data transformation
    ## Revise trainB, don't distinguish the Species
    data <- ddply(data,
                  .(Date, Block,
                    Year, Month, Week, Weekday
                    ),
                  summarize,
                  NumMosquitos = sum(NumMosquitos),
                  WnvPresent = as.integer(as.logical(sum(WnvPresent)))
                  )
    ## segment all data in the same year by the range of NumMosquitos
    year.data <- data[data$Year==year,]
```

```
    max.num <- tapply(year.data$NumMosquitos, year.data$Block, max)
    max.num <- data.frame(Block=names(max.num), MaxNumMosquitos=as.numeric(max.num))
    year.data <- merge(year.data, max.num)
    ## quantiles
    quant <- quantile(year.data[, group], seq(0, 1, 1/segments))
    ## starting and ending values of the quantile loop
    start.quant <- quant[-length(quant)]
    end.quant <- quant[-1]
    for (i in 1:length(start.quant)) {
        sq <- start.quant[i]
        eq <- end.quant[i]
        plot.year.data <- year.data[year.data$MaxNumMosquitos>=sq & year.data$MaxNumMosquitos<eq,]
        plot.year.data <- plot.year.data[order(plot.year.data$Block, plot.year.data$Date),]
        g <- ggplot(data = plot.year.data, aes(x = Date, y = NumMosquitos, colour=Block)) + geom_line()
        print(g)
    }
}
```

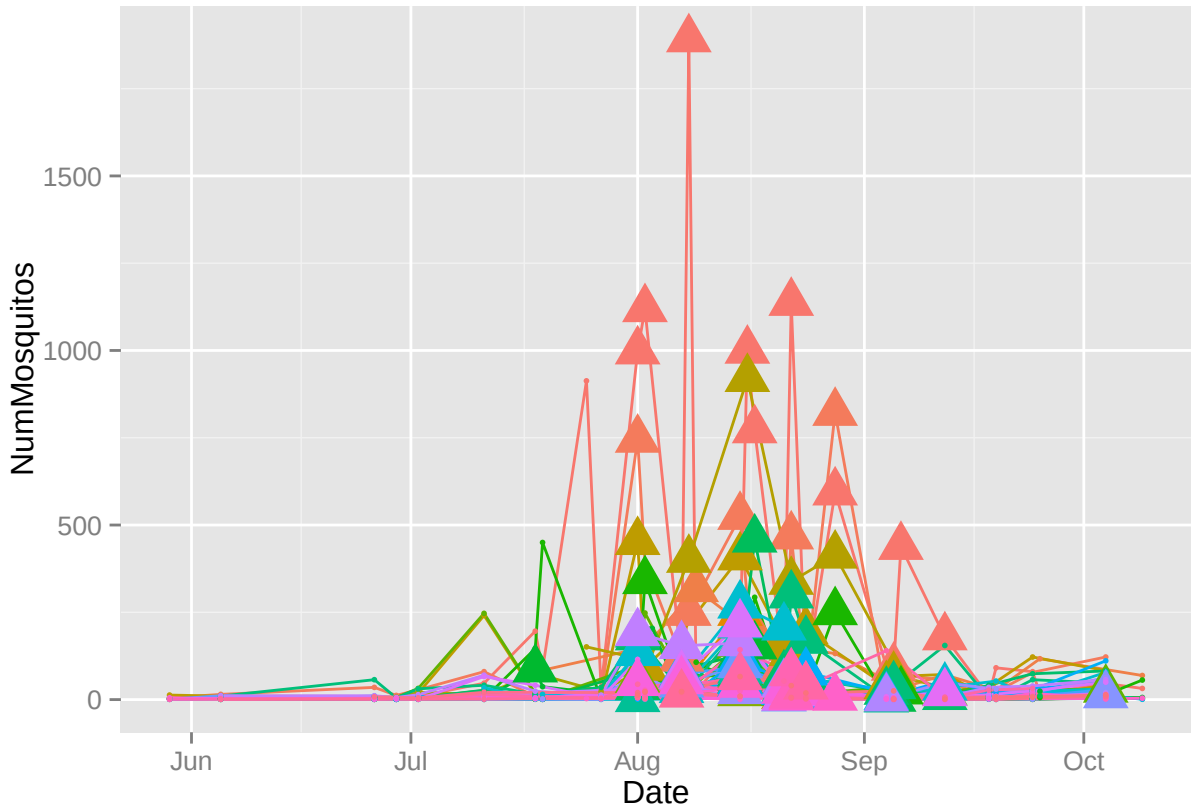
NumMosquitos for each year grouped by 4 ranges and blocks

```
plot.ts.group(trainB, 4, "MaxNumMosquitos", 2007)
plot.ts.group(trainB, 4, "MaxNumMosquitos", 2009)
plot.ts.group(trainB, 4, "MaxNumMosquitos", 2011)
plot.ts.group(trainB, 4, "MaxNumMosquitos", 2013)
```

NumMosquitos for each year grouped by blocks

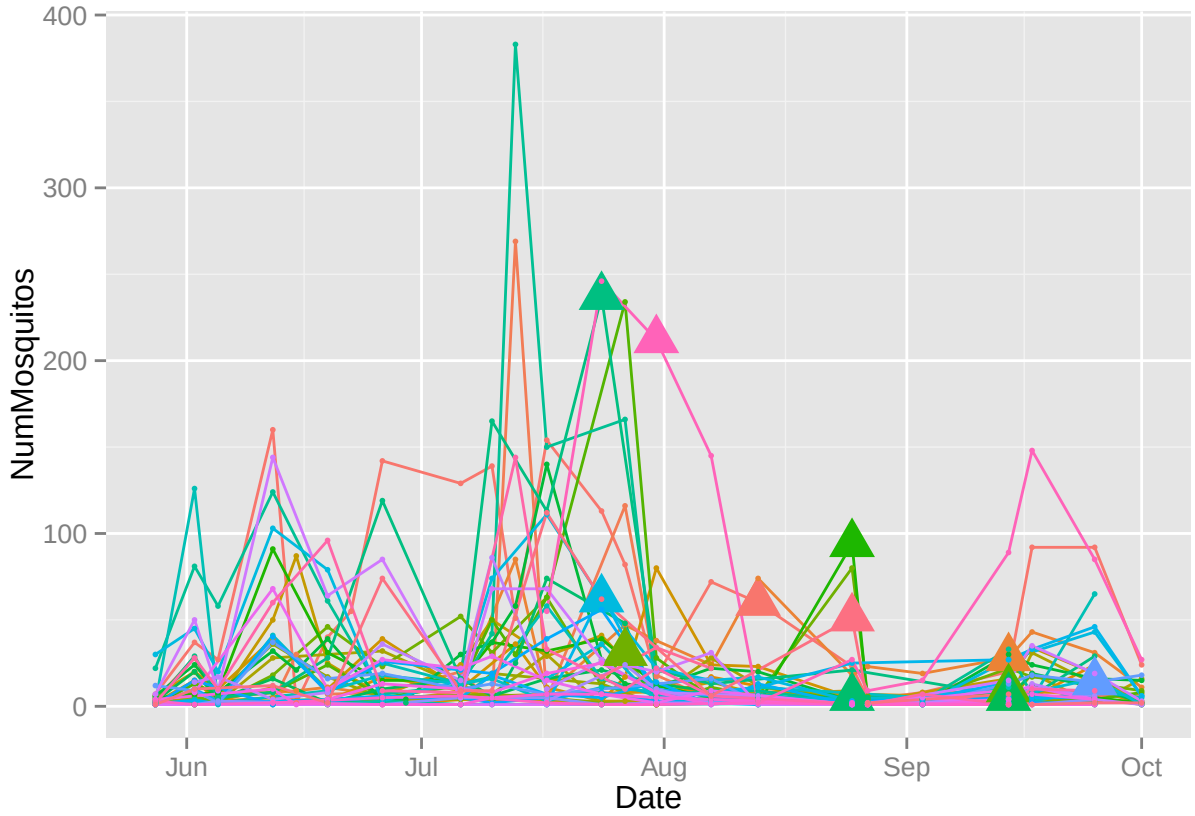
```
plot.ts.group(trainB, 1, "MaxNumMosquitos", 2007)
```

```
## year 2007 seg. 1
```



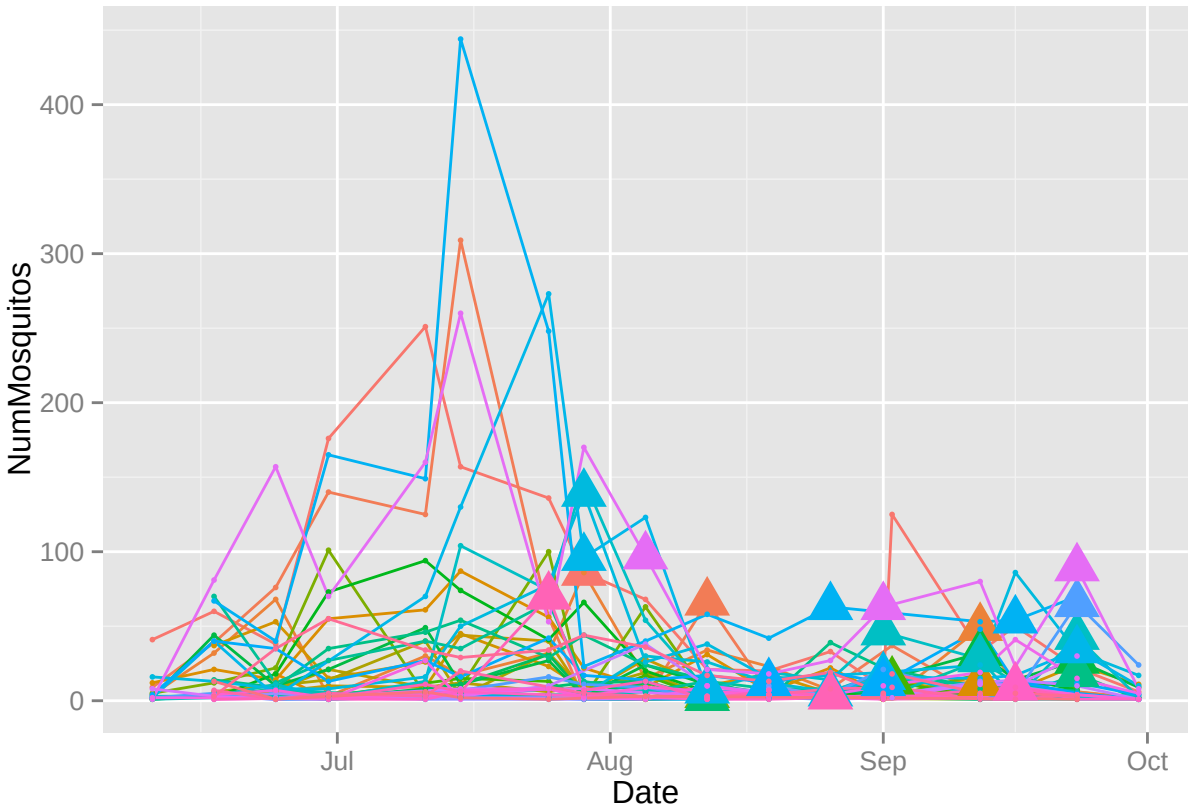

```
plot.ts.group(trainB, 1, "MaxNumMosquitos", 2009)
```

```
## year 2009 seg. 1
```



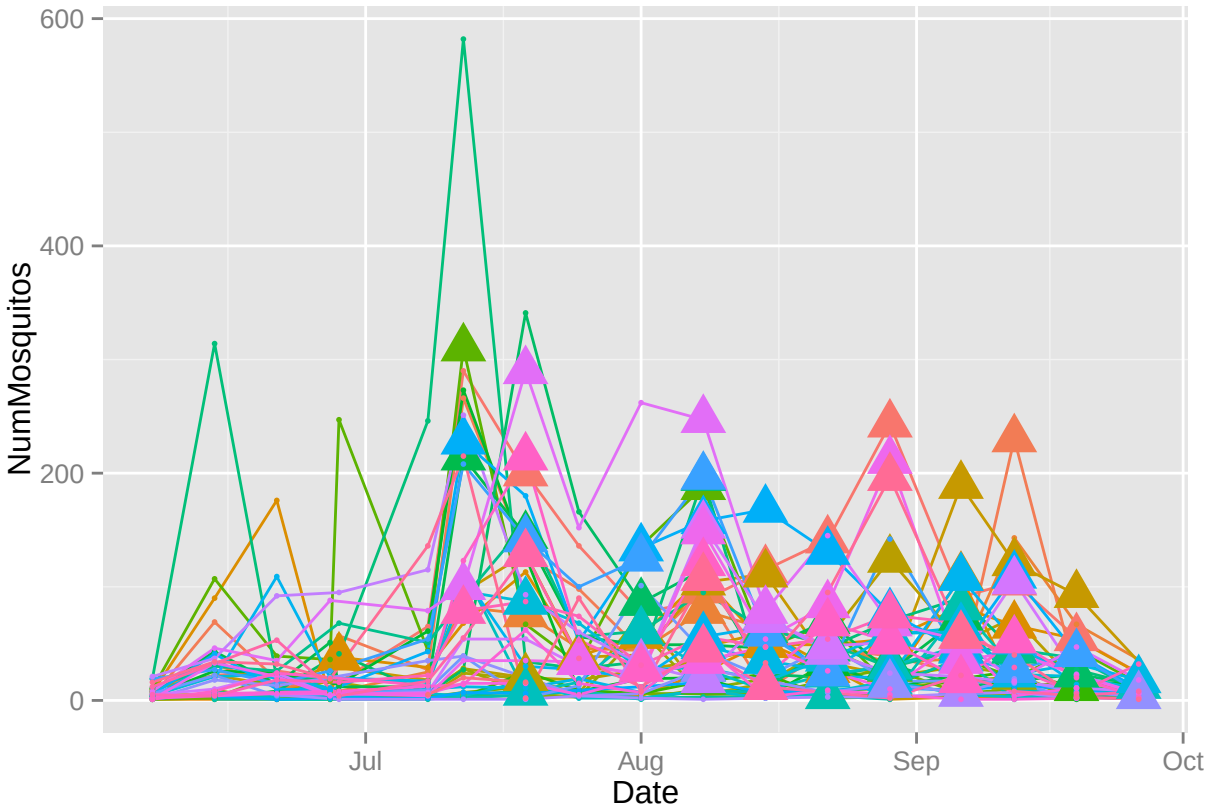

```
plot.ts.group(trainB, 1, "MaxNumMosquitos", 2011)
```

```
## year 2011 seg. 1
```




```
plot.ts.group(trainB, 1, "MaxNumMosquitos", 2013)
```

```
## year 2013 seg. 1
```



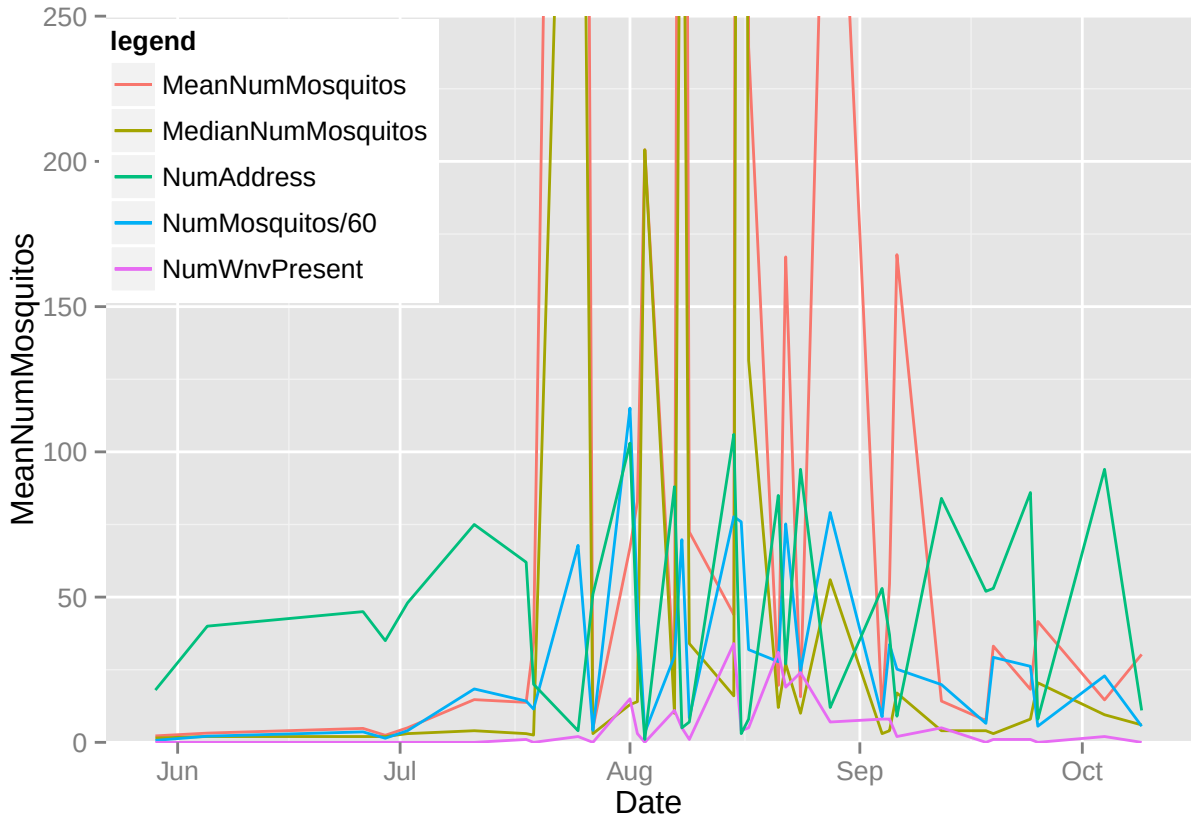
Total NumMosquitos versus Weather

```
## define the plotting function
library(ggplot2)
library(plyr)
plot.ts.mosquitos <- function(data, year, ylimit) {
    ## Revise trainB, don't distinguish the Species
    data <- ddply(data,
                  .(Date,
                    Year, Month, Week, Weekday,
                    Address
                    ),
                  summarize,
                  NumMosquitos = sum(NumMosquitos),
                  NumWnvPresent = sum(WnvPresent)
                  )
    data <- ddply(data,
                  .(Date,
                    Year, Month, Week, Weekday
                    ),
                  summarize,
                  MeanNumMosquitos = mean(NumMosquitos),
                  MedianNumMosquitos = median(NumMosquitos),
                  NumMosquitos = sum(NumMosquitos),
                  NumWnvPresent = sum(NumWnvPresent),
                  NumAddress = length(unique(Address))
```

```
                  )
    ## segment all data in the same year
    year.data <- data[data$Year==year,]
    g <- ggplot(data = year.data) + geom_line(aes(x=Date, y=MeanNumMosquitos, color="MeanNumMosquitos"))
    print(g)
}
```

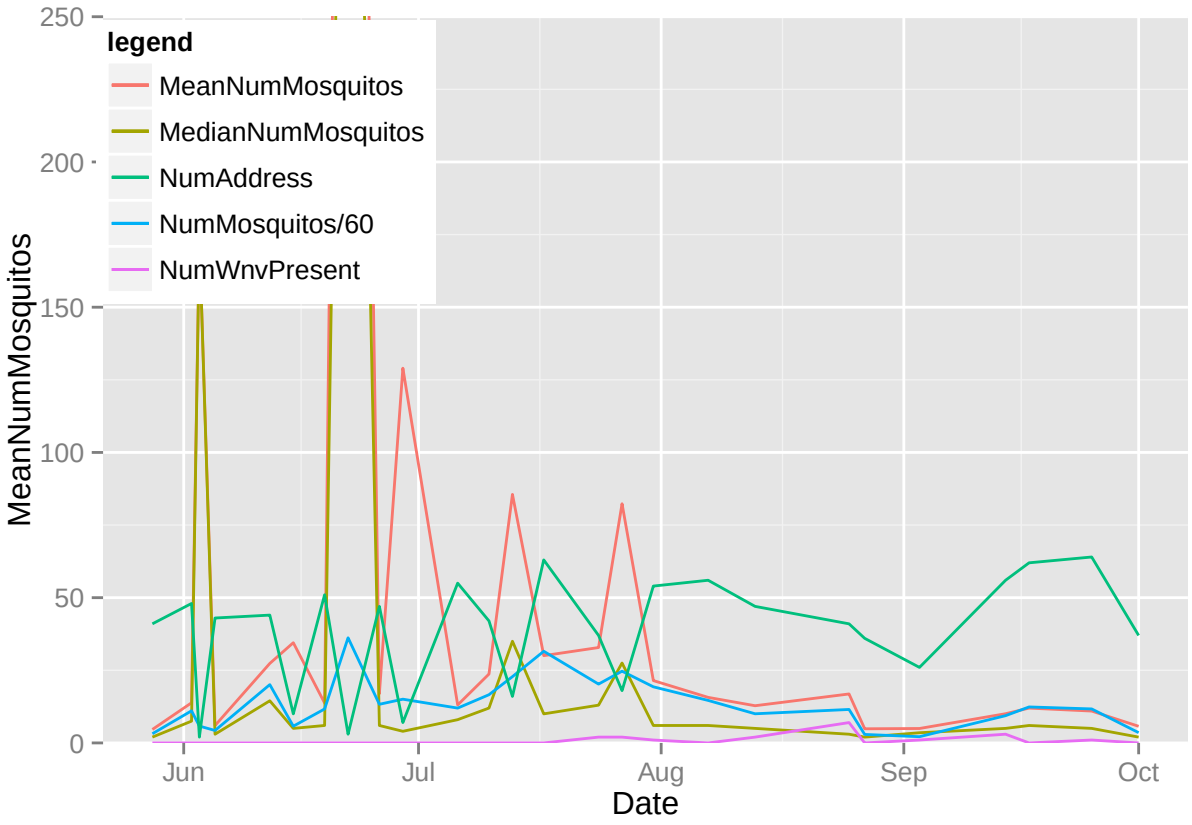
```
plot.ts.mosquitos(trainB, 2007, 250)
```

```
## year 2007
```



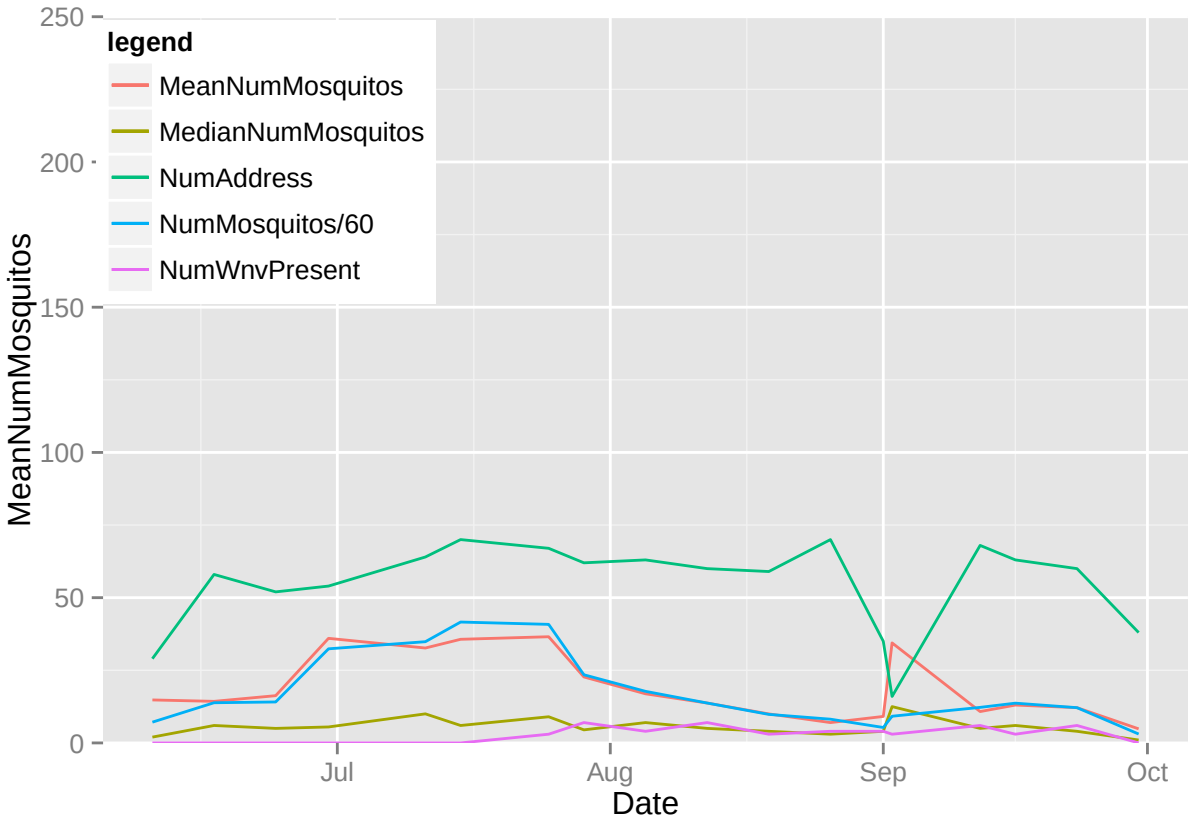

```
plot.ts.mosquitos(trainB, 2009, 250)
```

```
## year 2009
```



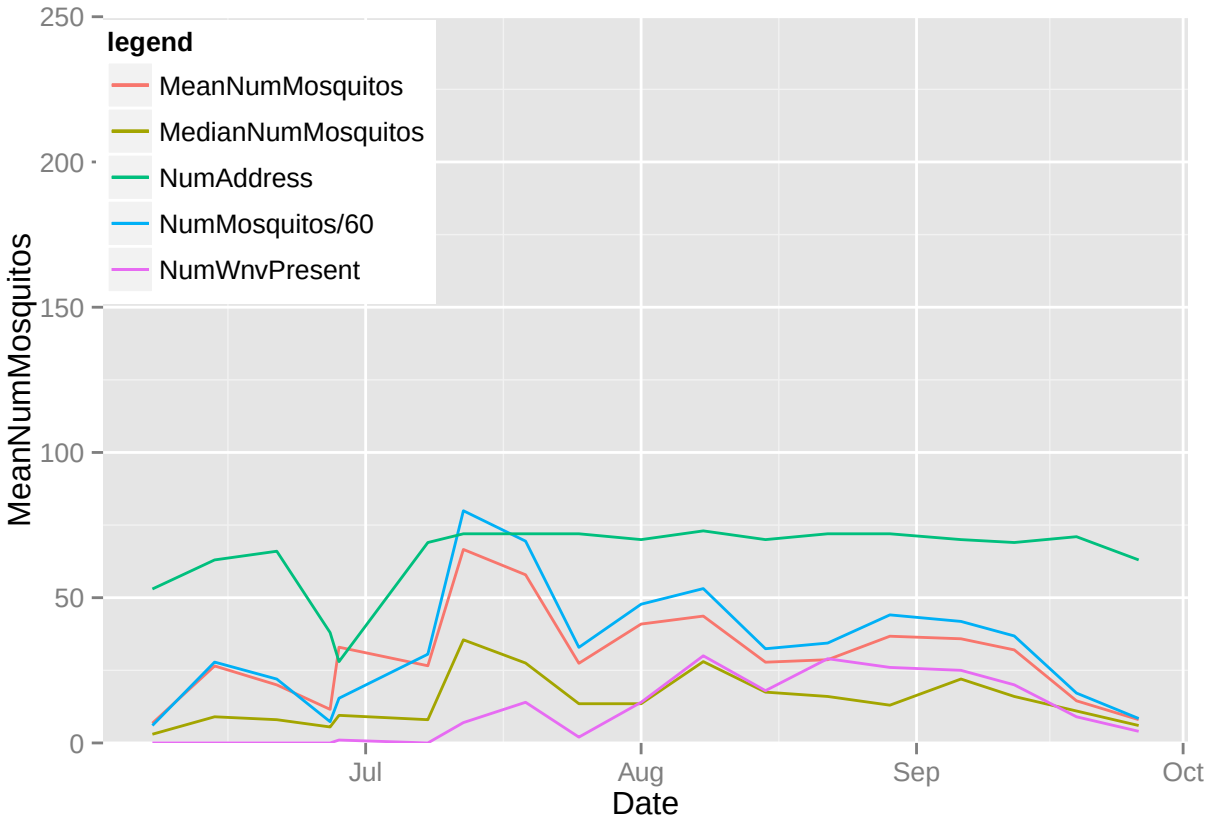

```
plot.ts.mosquitos(trainB, 2011, 250)
```

```
## year 2011
```




```
plot.ts.mosquitos(trainB, 2013, 250)
```

```
## year 2013
```

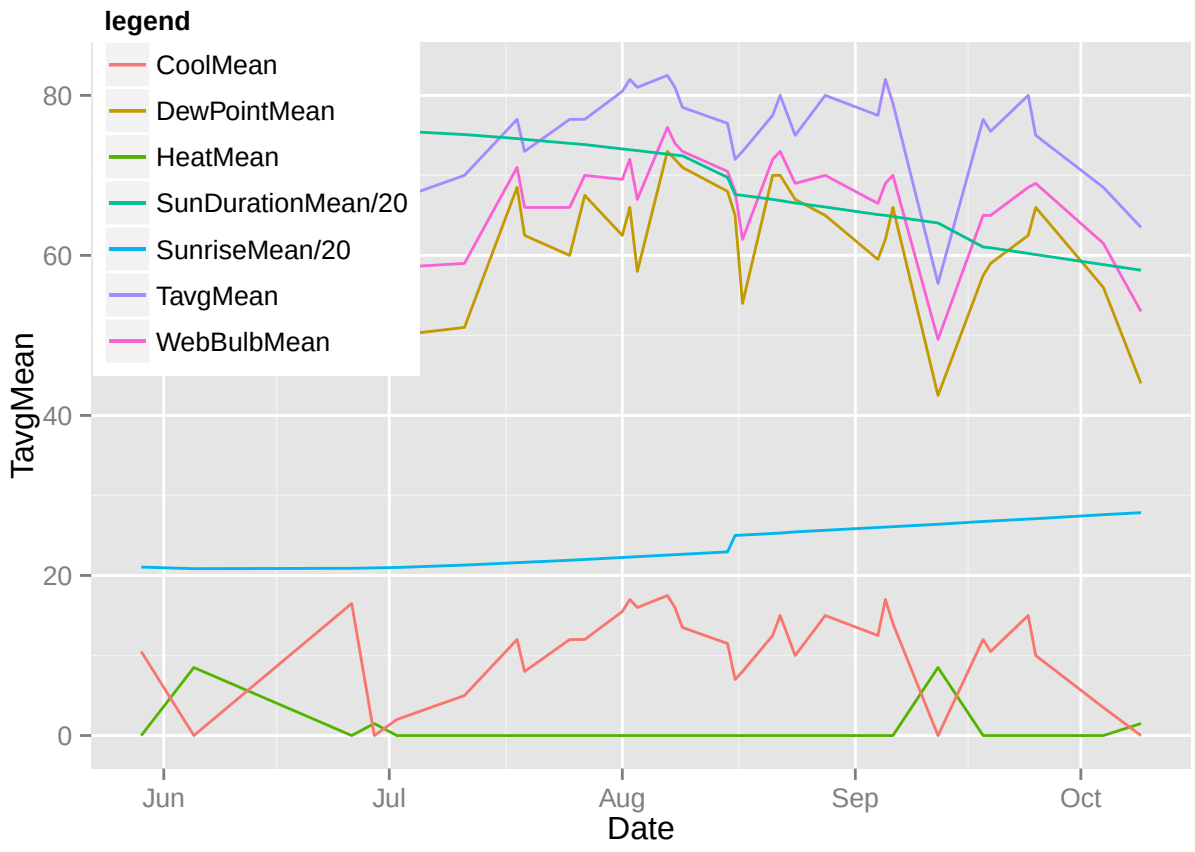


Weather in the four years

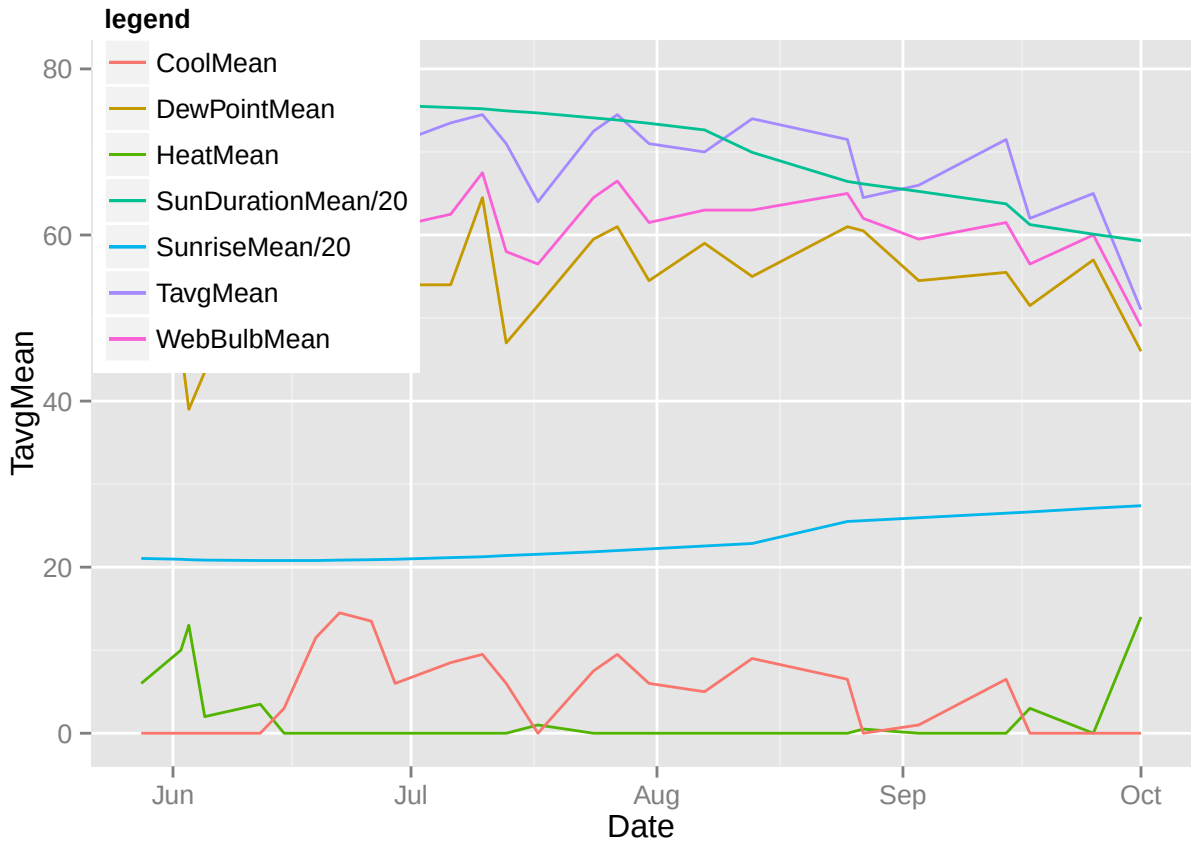
```
## define the plotting function
library(ggplot2)
library(plyr)
plot.ts.weather <- function(data, year) {
    ## Data transformation
    ## Revise trainB, don't distinguish the Species
    data <- ddply(data,
                  .(Date,
                    Year, Month, Week, Weekday
                    ),
                  summarize,
                  TavgMean = mean(unique(Tavg)),
                  DewPointMean = mean(unique(DewPoint)),
                  WetBulbMean = mean(unique(WetBulb)),
                  HeatMean = mean(unique(Heat)),
                  CoolMean = mean(unique(Cool)),
                  SunriseMean = mean(unique(Sunrise)),
                  SunsetMean = mean(unique(Sunset)),
                  SunDurationMean = mean(unique(Sunset)) - mean(unique(Sunrise))
                  )
    ## Scale the weather columns
    ## time.vars <- c("Date", "Year", "Month", "Week", "Weekday")
    ## scale.vars <- setdiff(names(data), time.vars)
    ## data <- data.frame(data[,time.vars], sapply(data[,scale.vars], scale))
```

```
    ## segment all data in the same year
    year.data <- data[data$Year==year,]
    g <- ggplot(data = year.data) + geom_line(aes(x=Date, y=TavgMean, color="TavgMean")) + geom_line(aes
        labs(color="legend") + theme(legend.position=c(0.15,0.8))
    print(g)
    }
```

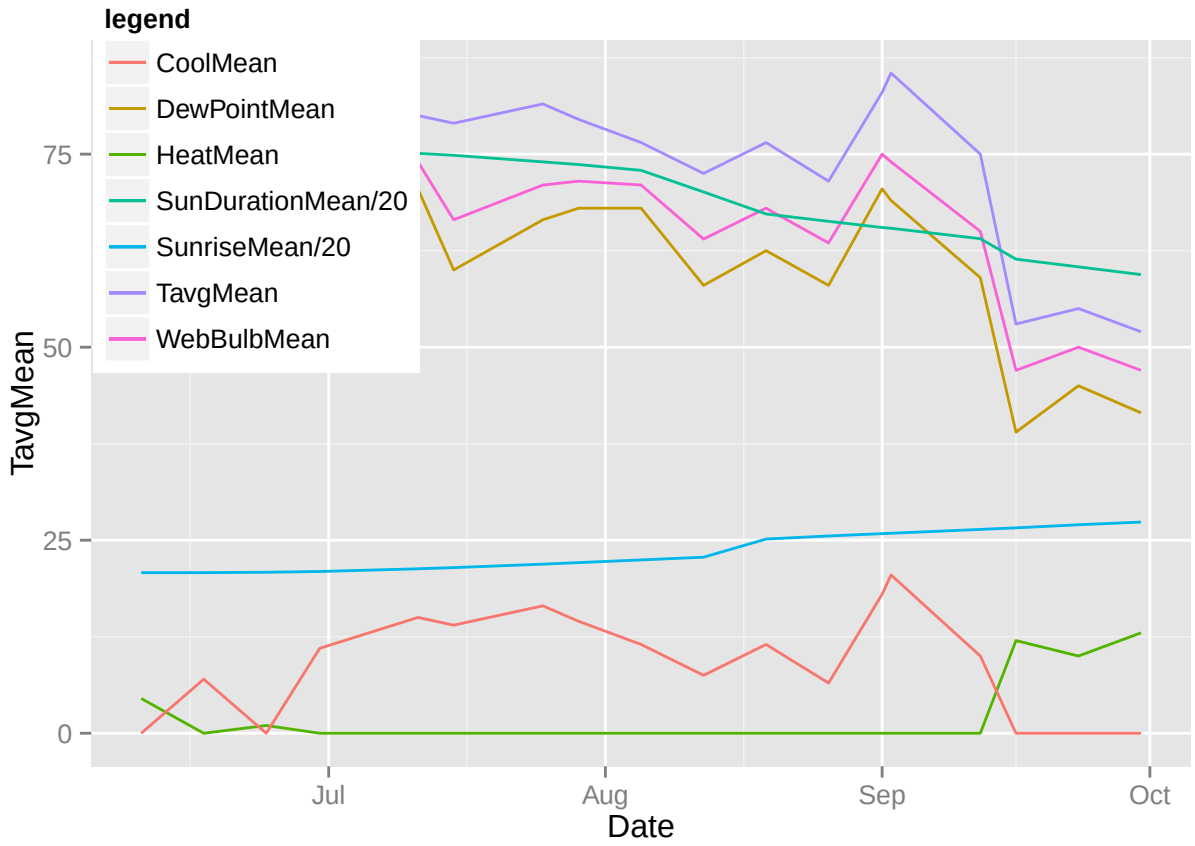
```
plot.ts.weather(trainB, 2007)
```



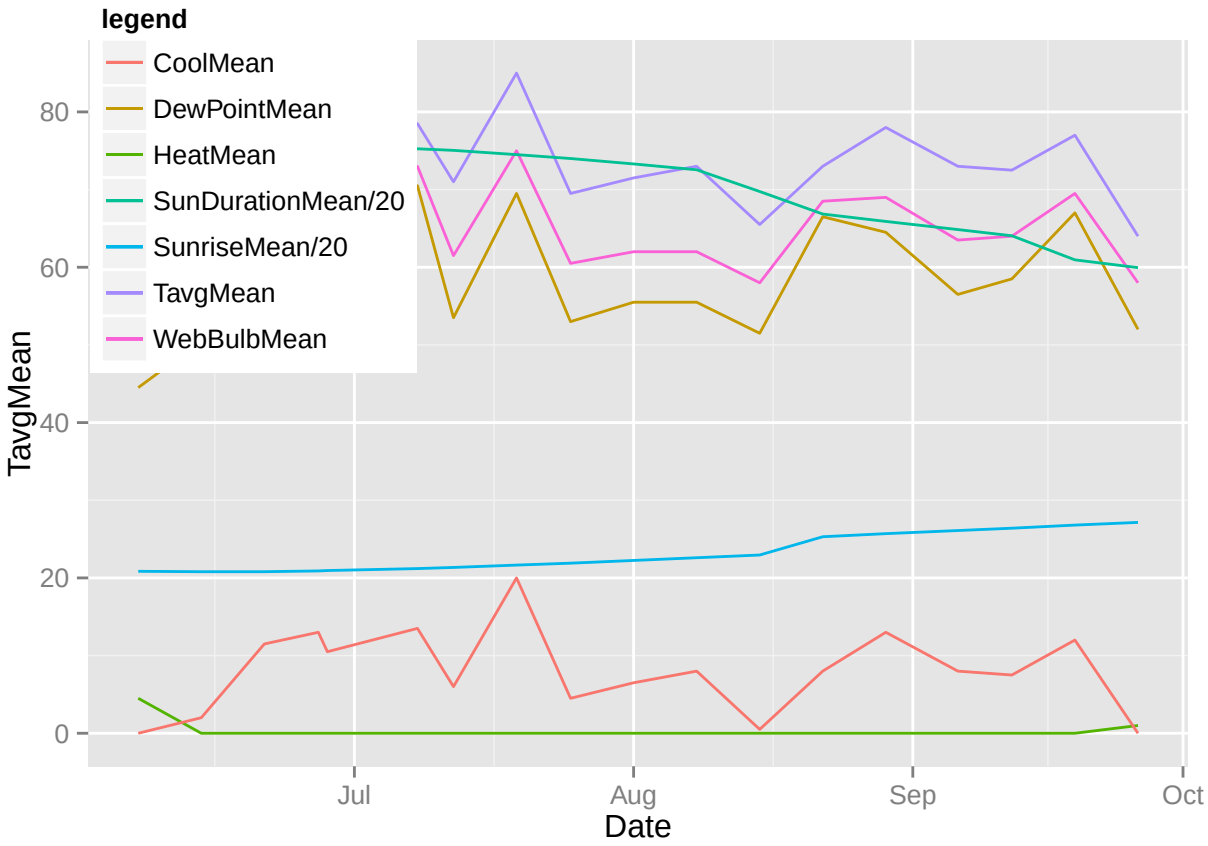

```
plot.ts.weather(trainB, 2009)
```




```
plot.ts.weather(trainB, 2011)
```




```
plot.ts.weather(trainB, 2013)
```


legend
CoolMean
DewPointMean
HeatMean
SunDurationMean/20
SunriseMean/20
TavgMean
WebBulbMean
TavgMean
80
60
40
20
0
Jul
Aug
Sep
Oct
Date