rain_gage.R

Yangsu

Sun Oct 18 01:07:27 2015

```
library(stringr)
theFiles<-dir("C:/Users/Yangsu/Desktop/rain data/",pattern="\\.txt")
theFiles
```

```
##  [1] "L0001.txt" "L0002.txt" "L0003.txt" "L0004.txt" "L0005.txt"
##  [6] "L0006.txt" "L0007.txt" "L0008.txt" "L0009.txt" "L0010.txt"
## [11] "L0011.txt" "L0012.txt" "L0101.txt" "L0102.txt" "L0103.txt"
## [16] "L0104.txt" "L0105.txt" "L0106.txt" "L0107.txt" "L0108.txt"
## [21] "L0109.txt" "L0110.txt" "L0111.txt" "L0112.txt" "L0201.txt"
## [26] "L0202.txt" "L0203.txt" "L0204.txt" "L0205.txt" "L0206.txt"
## [31] "L0207.txt" "L0208.txt" "L0209.txt" "L0210.txt" "L0211.txt"
## [36] "L0212.txt" "L0301.txt" "L0302.txt" "L0303.txt" "L0304.txt"
## [41] "L0305.txt" "L0306.txt" "L0307.txt" "L0308.txt" "L0309.txt"
## [46] "L0310.txt" "L0311.txt" "L0312.txt" "L0401.txt" "L0402.txt"
## [51] "L0403.txt" "L0404.txt" "L0405.txt" "L0406.txt" "L0407.txt"
## [56] "L0408.txt" "L0409.txt" "L0410.txt" "L0411.txt" "L0412.txt"
```

```
for (a in theFiles){
  nameToUse<-str_sub(string=a,start=1,end=7)
  temp<-read.csv(file=file.path("C:/Users/Yangsu/Desktop/rain data",a),
skip=2,stringsAsFactors = F)
  assign(x=nameToUse,value=temp)
}

rain<-rbind(L0001.t,    L0002.t,    L0003.t,    L0004.t,    L0005.t,
 L0006.t,    L0007.t,    L0008.t,    L0009.t,
          L0010.t,  L0011.t,    L0012.t,  L0101.t,  L0102.t,    L0103.t,
    L0104.t,    L0105.t,    L0106.t,
          L0107.t,  L0108.t,    L0109.t,    L0110.t,    L0111.t,    L01
12.t,  L0201.t,  L0202.t,    L0203.t,
          L0204.t,  L0205.t,    L0206.t,    L0207.t,    L0208.t,    L02
09.t,    L0210.t,    L0211.t,    L0212.t,
          L0301.t,  L0302.t,    L0303.t,    L0304.t,    L0305.t,    L03
06.t,    L0307.t,    L0308.t,    L0309.t,
          L0310.t,  L0311.t,    L0312.t,  L0401.t,  L0402.t,    L0403.t,
    L0404.t,    L0405.t,    L0406.t,
          L0407.t,  L0408.t,    L0409.t,    L0410.t,    L0411.t,    L04
12.t)

dim(rain)
```

```
## [1] 1827   25
```

```r
colnames(rain) <- 0:24
head(rain)
```

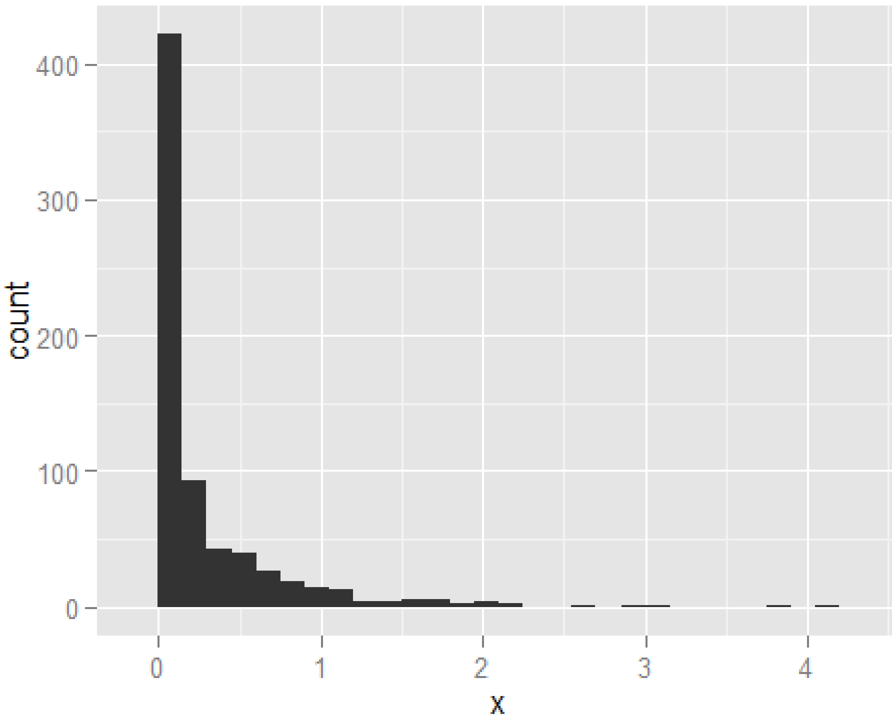
```
##    0    1    2    3    4    5    6    7    8    9    10   11   12   13
   14
## 1 1 ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
 ----
## 2 2 ---- ---- ---- ---- ---- ---- ---- ---- ---- ----  .03 T    ----
 ----
## 3 3 ---- ---- ---- T    ---- ---- ---- ---- ---- ---- ---- ---- ----
 ----
## 4 4 ----  .01 T    T    T    T    ---- ---- ----  .01 T    T    T
 T
## 5 5  .13  .07  .03 ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
 ----
## 6 6 ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
 ----
##     15   16   17   18   19   20   21   22   23   24
## 1 ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
## 2 ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
## 3 ---- ---- T    T     .02  .01 T    T    T    ----
## 4 ---- ---- T    T     .02  .03  .12  .21  .16   .2
## 5 ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
## 6 ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
```

```r
## to calculate each rain gage data during a rain storm, we add up numb
ers between two "----"
## "----" means no raining
## "T   " means a little rain but uncountable, so if T is between two n
umbers, then it is included in one storm
## I set "T   " to be the minor number so that it helps me with later c
alculation
## for "M   " I do not know what it means but it is always between two
"----" so I set it to be 0

rain[rain=="----"] <- 0
rain[rain=="M   "] <- 0
rain[rain=="M"] <- 0
rain[rain=="T   "] <- 10^(-8)
head(rain)
```

```
##    0   1   2     3     4     5     6 7 8 9  10    11    12    13    1
4 15
## 1 1   0   0     0     0     0     0 0 0 0   0     0     0     0
0  0
## 2 2   0   0     0     0     0     0 0 0 0   0   .03 1e-08     0
0  0
## 3 3   0   0     0 1e-08     0     0 0 0 0   0     0     0     0
0  0
## 4 4   0 .01 1e-08 1e-08 1e-08 1e-08 0 0 0 .01 1e-08 1e-08 1e-08 1e-0
8  0
```

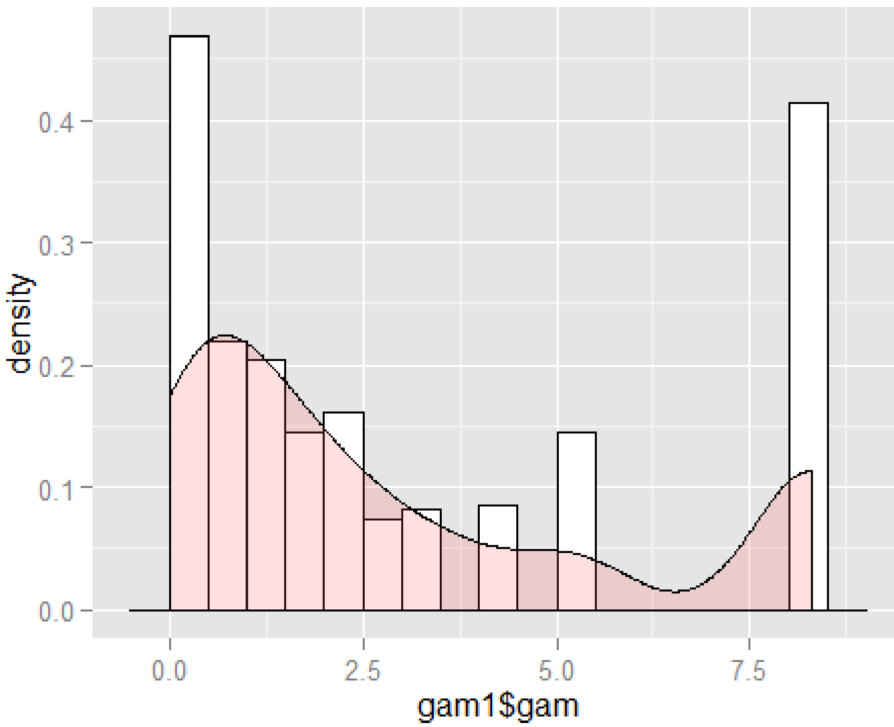
```
## 5 5 .13 .07   .03     0     0     0 0 0 0   0     0     0     0
0  0
## 6 6   0   0     0     0     0     0 0 0 0   0     0     0     0
0  0
##   16    17    18  19  20    21    22    23 24
## 1  0     0     0   0   0     0     0     0  0
## 2  0     0     0   0   0     0     0     0  0
## 3  0 1e-08 1e-08 .02 .01 1e-08 1e-08 1e-08  0
## 4  0 1e-08 1e-08 .02 .03   .12   .21   .16 .2
## 5  0     0     0   0   0     0     0     0  0
## 6  0     0     0   0   0     0     0     0  0

## delete the first column
r01<-rain[,(2:25)]
head(r01)

##      1   2     3     4     5     6 7 8 9  10    11    12    13    14
15 16
## 1   0   0     0     0     0     0 0 0 0   0     0     0     0     0
 0  0
## 2   0   0     0     0     0     0 0 0 0   0   .03 1e-08     0     0
 0  0
## 3   0   0     0 1e-08     0     0 0 0 0   0     0     0     0     0
 0  0
## 4   0 .01 1e-08 1e-08 1e-08 1e-08 0 0 0 .01 1e-08 1e-08 1e-08 1e-08
 0  0
## 5 .13 .07   .03     0     0     0 0 0 0   0     0     0     0     0
 0  0
## 6   0   0     0     0     0     0 0 0 0   0     0     0     0     0
 0  0
##      17    18  19  20    21    22    23 24
## 1     0     0   0   0     0     0     0  0
## 2     0     0   0   0     0     0     0  0
## 3 1e-08 1e-08 .02 .01 1e-08 1e-08 1e-08  0
## 4 1e-08 1e-08 .02 .03   .12   .21   .16 .2
## 5     0     0   0   0     0     0     0  0
## 6     0     0   0   0     0     0     0  0

## change class of character to class of numeric
bos <- as.data.frame(sapply(r01, as.numeric))
bosrain<-bos[complete.cases(bos), ]
View(bosrain)

## change data frame into vector
brain <- as.vector(t(bosrain))

## use function to get rain data for each storm
## we also build up a vector to put the results in

sum <- 0
```

```
j=1
vector<-0
for(i in 1:length(brain))
{
  if(brain[i] != 0)
  {
    sum=sum+brain[i]
  }
  if(brain[i]==0 && sum!=0)
  {
    vector[j]=sum
    j=j+1
    sum=0
  }
  if(brain[i]!=0 & i==length(brain))
  {
    vector[j]=sum
  }
}

## in order to delete those T without surrounding by numbers, we choose
 to keep only two digits parts
vector1<-round(vector, 2)
v2<-vector1[vector1 != 0.00]

## so that v2 is boston logan airport rain gage data which looks the sa
me as illinois rain data
## next we start with rain gage distribution analysis
class(v2)

## [1] "numeric"

logan <- data.frame(v2)
colnames(logan) <- "x"
library(ggplot2)
qplot(x, data=logan, geom = "histogram",binwidth=.15)
```

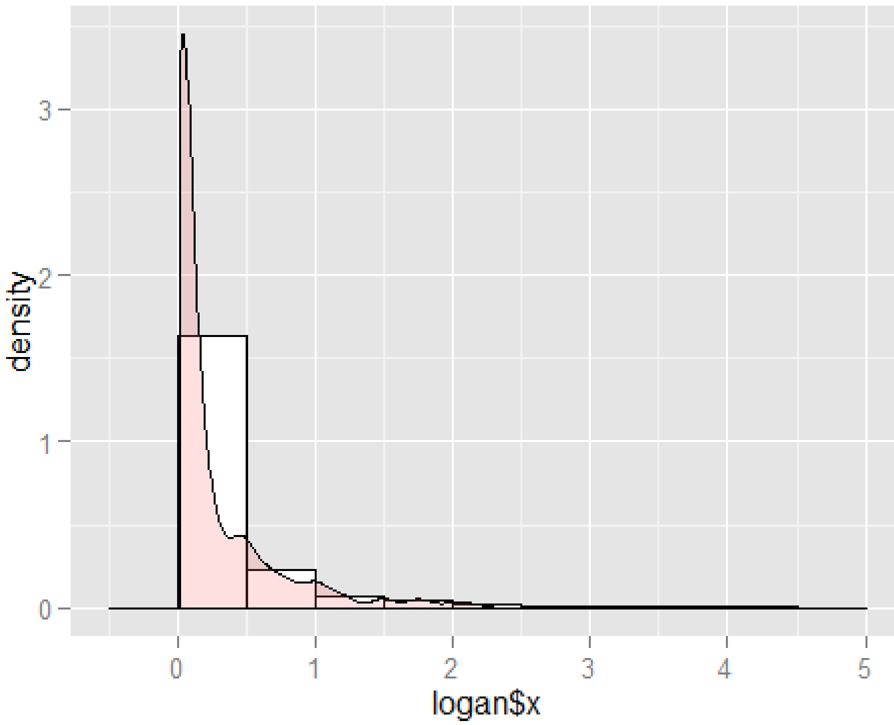



```
## it looks like gamma distribution

mean(logan$x)

## [1] 0.2831108

var(logan$x)

## [1] 0.2218382

alpha <- mean(logan$x)^2/var(logan$x)  # alpha = 0.36
lambda <- mean(logan$x)/var(logan$x)   # lambda = 1.28


gam<-(lambda^(alpha)/gamma(alpha))*(logan$x^(alpha-1))*exp(-lambda*loga
n$x)
gam1<-data.frame(gam)


## gamma distribution density plot
library(ggplot2)
ggplot(gam1, aes(x=gam1$gam)) +
  geom_histogram(aes(y=..density..),      # Histogram with density inst
ead of count on y-axis
                 binwidth=.5,
                 colour="black", fill="white") +
```

```
  geom_density(alpha=.2, fill="#FF6666")  # Overlay with transparent de
nsity plot
```




```
## data distribution density plot
ggplot(logan, aes(x=logan$x)) +
  geom_histogram(aes(y=..density..),      # Histogram with density inst
ead of count on y-axis
                 binwidth=.5,
                 colour="black", fill="white") +
  geom_density(alpha=.2, fill="#FF6666")  # Overlay with transparent de
nsity plot
```




```
# for Variance & confidence interval

## using MEM
lam<-mean(logan$x)/(sd(logan$x)^2)
alp<-(mean(logan$x))^2/(sd(logan$x)^2)

B<-1000
Tboot1<-rep(0,B)
Tboot2<-rep(0,B)
for(i in 1:B){
  x <- sample(logan$x,1000,replace=TRUE)
  Tboot1[i] <- mean(x)/(sd(x)^2)
  Tboot2[i] <- (mean(x))^2/(sd(x)^2)
}

Percentile1 <- c(quantile(Tboot1,.025),quantile(Tboot1,.975))
pivotal1 <- c((2*lam - quantile(Tboot1, .975)),(2*lam - quantile(Tboot1,
 .025)))

cat("Method       95% Interval\n")

## Method       95% Interval

cat("Pivotal1     (", pivotal1[1], ",     ", pivotal1[2], ") \n")

## Pivotal1     ( 0.9967536 ,      1.483044 )
```

```r
cat("Percentile1  (", Percentile1[1], ",    ", Percentile1[2], ") \n")

## Percentile1  ( 1.069364 ,     1.555654 )

Percentile2 <- c(quantile(Tboot2,.025),quantile(Tboot2,.975))
pivotal2 <- c((2*alp - quantile(Tboot2, .975)),(2*alp - quantile(Tboot2,
 .025)))

cat("Method       95% Interval\n")

## Method       95% Interval

cat("Pivotal2     (", pivotal2[1], ",     ", pivotal2[2], ") \n")

## Pivotal2     ( 0.2965933 ,      0.4140251 )

cat("Percentile2  (", Percentile2[1], ",    ", Percentile2[2], ") \n")

## Percentile2  ( 0.308589 ,     0.4260208 )

##  for MLE method

  mle.x <- logan$x
  n <- length(logan$x)
  # first we need to have alpha and lambda from MEM
  mem.alp <- mean(mle.x)^2/var(mle.x)
  mem.lam <- (mean(mle.x))/var(mle.x)
  mem.alp

## [1] 0.3613071

  mem.lam

## [1] 1.276204

  # second we use MLE to get parameter value
  minus.likelihood <- function(theta) {-(n*theta[1]*log(theta[2])-n*lga
mma(theta[1])+(theta[1]-1)*sum(log(mle.x))-theta[2]*sum(mle.x))}

  max.likelihood <- nlminb(start=c(mem.alp, mem.lam), obj = minus.likel
ihood)

  max.likelihood$par

## [1] 0.5461541 1.9291179
```