MemoryEffi.R

arcs

Mon Nov 27 13:55:08 2017

```
library(ggplot2)
library(scales)

setwd("/home/arcs/Oct14/DataCSV")
getwd()
```

```
## [1] "/home/arcs/Oct14/DataCSV"
```

```
newdata <- read.csv("14Oct2017EfficiencyMem1.csv", header = T, sep=",")

########################################################################
############# Studying the structure of Data ###########################
########################################################################
names(newdata)
```

```
##  [1] "RemoteWallClockTime"      "ExitBySignal"
##  [3] "ExitCode"                 "ExitSignal"
##  [5] "ExitStatus"               "RemoteSysCpu"
##  [7] "RemoteUserCpu"            "CumulativeSuspensionTime"
##  [9] "RequestMemory"            "MemoryUsage"
## [11] "default_maxMemory"        "maxMemory"
## [13] "CumulativeRemoteSysCpu"   "CumulativeRemoteUserCpu"
## [15] "Remote_JobUniverse"       "JobUniverse"
```

```
str(newdata)
```

```
## 'data.frame':    257561 obs. of  16 variables:
##  $ RemoteWallClockTime     : Factor w/ 34398 levels "0","1","10","100",..: 26685 1194 10337 31892 11
##  $ ExitBySignal            : Factor w/ 2 levels "False","True": 1 1 1 1 1 1 1 1 1 1 ...
##  $ ExitCode                : Factor w/ 7 levels "0","1","126",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ExitSignal              : Factor w/ 5 levels "1","11","15",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ ExitStatus              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RemoteSysCpu            : int  0 0 4 208 0 0 1 0 1 0 ...
##  $ RemoteUserCpu           : int  1 4 8 6486 3 4 4 4 3 4 ...
##  $ CumulativeSuspensionTime: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RequestMemory           : int  1900 1900 4000 2000 1900 1900 1900 1900 1900 1900 ...
##  $ MemoryUsage             : Factor w/ 95 levels "0","1","10","11",..: 24 39 63 47 39 39 39 39 39 39
##  $ default_maxMemory       : int  2130 2130 2130 2130 2130 2130 2130 2130 2130 2130 ...
##  $ maxMemory               : Factor w/ 10 levels "0","16000","18000",..: 4 4 8 5 4 4 4 4 4 4 ...
##  $ CumulativeRemoteSysCpu  : Factor w/ 3143 levels "0.0","1.0","10.0",..: 3143 3143 1692 791 3143 314
##  $ CumulativeRemoteUserCpu : Factor w/ 31043 levels "0.0","1.0","10.0",..: 2 19203 29995 28645 10275
##  $ Remote_JobUniverse      : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ JobUniverse             : int  5 5 5 5 5 5 5 5 5 5 ...
```

```
summary(newdata)
```

```
##  RemoteWallClockTime ExitBySignal   ExitCode      ExitSignal
##  None   : 61843      False:255905   0   :115272   1   :   867
##  1      : 23953      True :  1656   1   :   919   11  :     1
##  141    : 11093                     126 :   277   15  :    37
```

```
##  140    :  9962                    127 : 35548   9    :   752
##  2      :  9514                    137 : 42694   None:255904
##  142    :  7431                    3   :   238
##  (Other):133765                    None: 62613
##    ExitStatus  RemoteSysCpu      RemoteUserCpu
##  Min.   :0    Min.   :     0.0   Min.   :      0
##  1st Qu.:0    1st Qu.:     0.0   1st Qu.:      0
##  Median :0    Median :     0.0   Median :      4
##  Mean   :0    Mean   :   298.2   Mean   :  13123
##  3rd Qu.:0    3rd Qu.:     7.0   3rd Qu.:     18
##  Max.   :0    Max.   :113711.0   Max.   :1929221
##
##  CumulativeSuspensionTime RequestMemory    MemoryUsage
##  Min.   :0                Min.   :    0   27     :45698
##  1st Qu.:0                1st Qu.: 1900   0      :43474
##  Median :0                Median : 2130   7325   :27518
##  Mean   :0                Mean   : 3389   1709   :24345
##  3rd Qu.:0                3rd Qu.: 2130   None   :19839
##  Max.   :0                Max.   :18000   1954   : 8983
##                                           (Other):87704
##  default_maxMemory   maxMemory      CumulativeRemoteSysCpu
##  Min.   :2130      None   :101601   None   :103717
##  1st Qu.:2130      1900   : 54248   0.0    : 32566
##  Median :2130      0      : 33014   1.0    : 22333
##  Mean   :2130      4000   : 26913   3.0    : 11137
##  3rd Qu.:2130      16000  : 24143   4.0    :  8724
##  Max.   :2130      2000   : 13393   5.0    :  5855
##                    (Other):  4249   (Other): 73229
##  CumulativeRemoteUserCpu Remote_JobUniverse  JobUniverse
##  None   :70295           Min.   :5          Min.   :5
##  4.0    :33892           1st Qu.:5          1st Qu.:5
##  0.0    :32614           Median :5          Median :5
##  3.0    : 7485           Mean   :5          Mean   :5
##  11.0   : 7028           3rd Qu.:5          3rd Qu.:5
##  10.0   : 6416           Max.   :5          Max.   :5
##  (Other):99831
```

```
#########################################################################
############## Conversion to numeric values ##############################
#########################################################################
newdata[,"RemoteWallClockTime"] <- as.numeric(as.character(newdata[,"RemoteWallClockTime"])) #RemoteWal
```

```
## Warning: NAs introduced by coercion
```

```
newdata[, "ExitCode"] <- as.numeric(as.character(newdata[, "ExitCode"]))
```

```
## Warning: NAs introduced by coercion
```

```
newdata[, "MemoryUsage"] <- as.numeric(as.character(newdata[, "MemoryUsage"]))
```

```
## Warning: NAs introduced by coercion
```

```
#########################################################################
###################### Data Cleansing ###################################
#########################################################################

unique(newdata$JobUniverse)
```

```
## [1] 5
```

```
unique(newdata$Remote_JobUniverse)
```

```
## [1] 5
```

```
unique(newdata$ExitCode)
```

```
## [1]   0  NA   1 137   3 127 126
```

```
newdata2 <- subset(newdata, newdata$ExitCode == 0)
unique(newdata2$ExitCode)
```

```
## [1] 0
```

```
unique(newdata2$JobUniverse)
```

```
## [1] 5
```

```
unique(newdata2$Remote_JobUniverse)
```

```
## [1] 5
```

```
########################################################################
############### Computation of efficiency ##############################
########################################################################

newdata2$CPUTime <- newdata2$RemoteSysCpu + newdata2$RemoteUserCpu
newdata2$WallTime <- newdata2$RemoteWallClockTime - newdata2$CumulativeSuspensionTime
newdata2$Efficiency <- newdata2$CPUTime/ newdata2$WallTime



#Cleanseing data by removing NA rows
newdata2 <- subset(newdata2, newdata2$Efficiency != "NA")
#newdata3 <- subset(newdata2, select =  c(CPUTime, WallTime, Efficiency))
#newdata3 <- na.omit(newdata3)
summary(newdata2)
```

```
##  RemoteWallClockTime ExitBySignal      ExitCode ExitSignal      ExitStatus
##  Min.   :     0      False:113703   Min.   :0   1   :     0   Min.   :0
##  1st Qu.:    31      True :     0   1st Qu.:0   11  :     0   1st Qu.:0
##  Median :   140                     Median :0   15  :     0   Median :0
##  Mean   :  5493                     Mean   :0   9   :     0   Mean   :0
##  3rd Qu.:   144                     3rd Qu.:0   None:113703   3rd Qu.:0
##  Max.   :301559                     Max.   :0                 Max.   :0
##
##   RemoteSysCpu      RemoteUserCpu     CumulativeSuspensionTime
##  Min.   :     0.0   Min.   :      0   Min.   :0
##  1st Qu.:     0.0   1st Qu.:      4   1st Qu.:0
##  Median :     1.0   Median :      5   Median :0
##  Mean   :   475.7   Mean   :  16228   Mean   :0
##  3rd Qu.:     5.0   3rd Qu.:     11   3rd Qu.:0
##  Max.   :113711.0   Max.   :1929221   Max.   :0
##
##  RequestMemory    MemoryUsage      default_maxMemory   maxMemory
##  Min.   :    0   Min.   :    0.0   Min.   :2130      1900   :54248
##  1st Qu.: 1900   1st Qu.:   10.0   1st Qu.:2130      4000   :26909
```

```
##  Median : 2000   Median :   27.0   Median :2130      2000   :13392
##  Mean   : 3086   Mean   :  386.9   Mean   :2130      None   :11760
##  3rd Qu.: 4000   3rd Qu.:   27.0   3rd Qu.:2130      16000  : 3913
##  Max.   :18000   Max.   :19532.0   Max.   :2130      3700   : 1483
##                                                      (Other): 1998
##  CumulativeRemoteSysCpu CumulativeRemoteUserCpu Remote_JobUniverse
##  None   :33534          4.0    :33668           Min.   :5
##  1.0    :21455          3.0    : 7125           1st Qu.:5
##  3.0    :11131          11.0   : 7028           Median :5
##  4.0    : 8667          10.0   : 6416           Mean   :5
##  5.0    : 5809          6.0    : 5178           3rd Qu.:5
##  2.0    : 3735          1.0    : 4695           Max.   :5
##  (Other):29372          (Other):49593
##   JobUniverse    CPUTime           WallTime        Efficiency
##  Min.   :5    Min.   :      0   Min.   :     0   Min.   :0.00000
##  1st Qu.:5    1st Qu.:      4   1st Qu.:    31   1st Qu.:0.02857
##  Median :5    Median :      7   Median :   140   Median :0.14236
##  Mean   :5    Mean   :  16704   Mean   :  5493   Mean   :    Inf
##  3rd Qu.:5    3rd Qu.:     17   3rd Qu.:   144   3rd Qu.:0.50000
##  Max.   :5    Max.   :1941994   Max.   :301559   Max.   :    Inf
##                                                  NA's   :97
```

```
################################################################################
str(newdata2)
```

```
## 'data.frame':    113703 obs. of  19 variables:
##  $ RemoteWallClockTime     : num  53 139 33 6967 138 ...
##  $ ExitBySignal            : Factor w/ 2 levels "False","True": 1 1 1 1 1 1 1 1 1 1 ...
##  $ ExitCode                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ ExitSignal              : Factor w/ 5 levels "1","11","15",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ ExitStatus              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RemoteSysCpu            : int  0 0 4 208 0 0 1 0 1 0 ...
##  $ RemoteUserCpu           : int  1 4 8 6486 3 4 4 4 3 4 ...
##  $ CumulativeSuspensionTime: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RequestMemory           : int  1900 1900 4000 2000 1900 1900 1900 1900 1900 1900 ...
##  $ MemoryUsage             : num  2 27 4 318 27 27 27 27 27 27 ...
##  $ default_maxMemory       : int  2130 2130 2130 2130 2130 2130 2130 2130 2130 2130 ...
##  $ maxMemory               : Factor w/ 10 levels "0","16000","18000",..: 4 4 8 5 4 4 4 4 4 4 ...
##  $ CumulativeRemoteSysCpu  : Factor w/ 3143 levels "0.0","1.0","10.0",..: 3143 3143 1692 791 3143 314
##  $ CumulativeRemoteUserCpu : Factor w/ 31043 levels "0.0","1.0","10.0",..: 2 19203 29995 28645 10275
##  $ Remote_JobUniverse      : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ JobUniverse             : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ CPUTime                 : int  1 4 12 6694 3 4 5 4 4 4 ...
##  $ WallTime                : num  53 139 33 6967 138 ...
##  $ Efficiency              : num  0.0189 0.0288 0.3636 0.9608 0.0217 ...
```

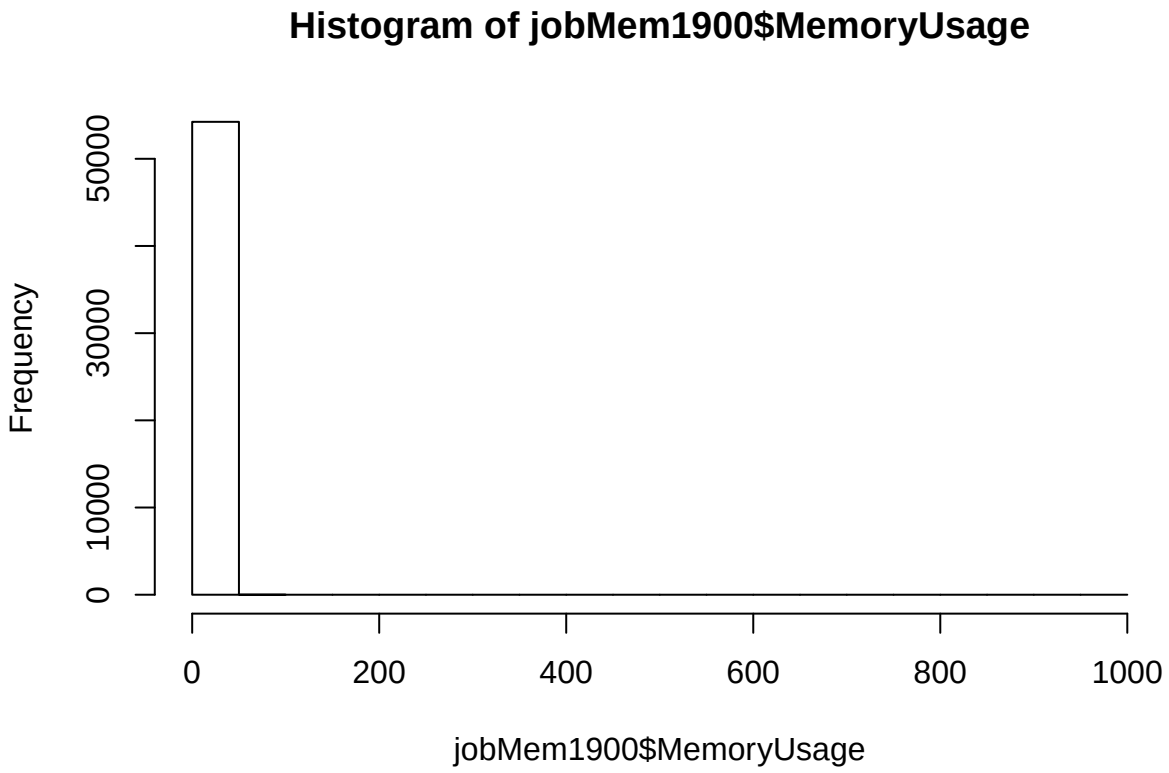
```
unique(newdata2$MemoryUsage)
```

```
##  [1]     2    27     4   318  1709  7325     3   440     0    15  9766
## [12] 14649    13   733    10     8    32  1221  2930   245    25  2686
## [23]    20  1954   977 12208  1465  2442    49    30   489   464  2198
## [34]    22   342   269    18     5    98   391   220    42   416   123
## [45]     1   367  3174   293    35   171  4639    37 17090  3418   147
## [56]    40  3907    74  3663    44    47   196 19532  4883  4151
```

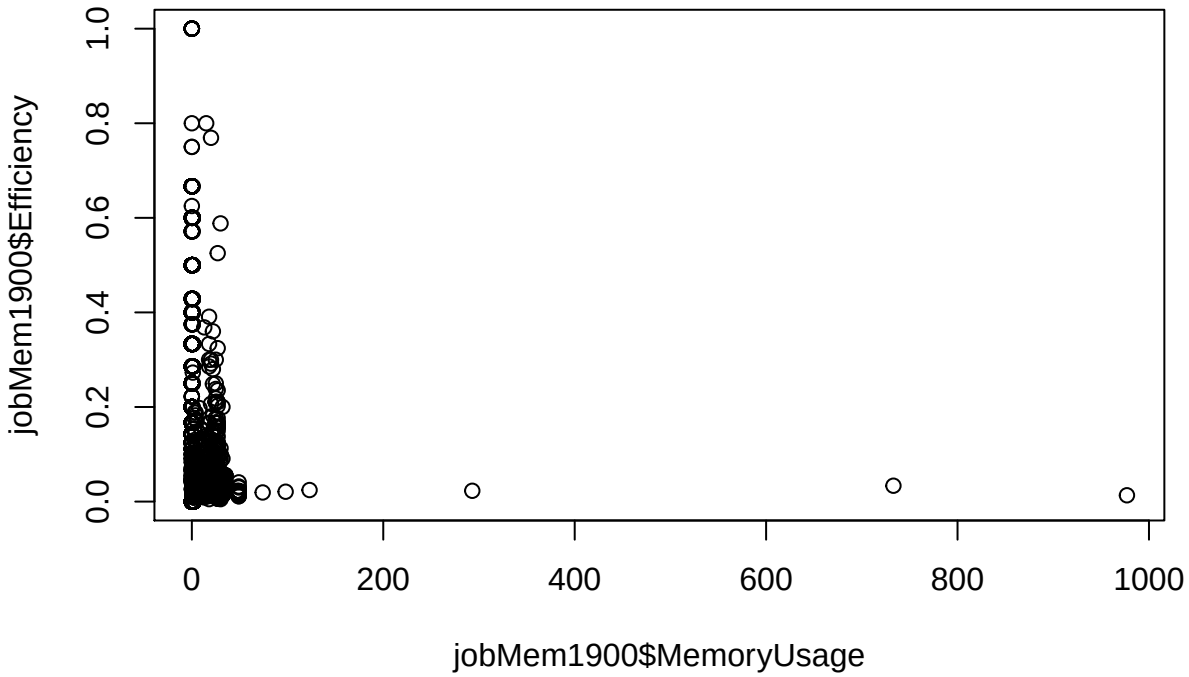
```r
unique(newdata2$RequestMemory)
```

```
##  [1]  1900  4000  2000 16000  3700  2130  3200     0 18000  8000
```

```r
#################### Jobs with memory request 1900 ########################################
jobMem1900 <- subset(newdata2, newdata2$RequestMemory == 1900)
hist(jobMem1900$MemoryUsage)
```




```r
plot(jobMem1900$MemoryUsage, jobMem1900$Efficiency)
```




```
TotalCPUTime_JobMem1900 <- sum(as.numeric(jobMem1900$CPUTime))
TotalWallTime_JobMem1900 <- sum(jobMem1900$WallTime)
TotalCPUTime_JobMem1900
```

```
## [1] 230546
```

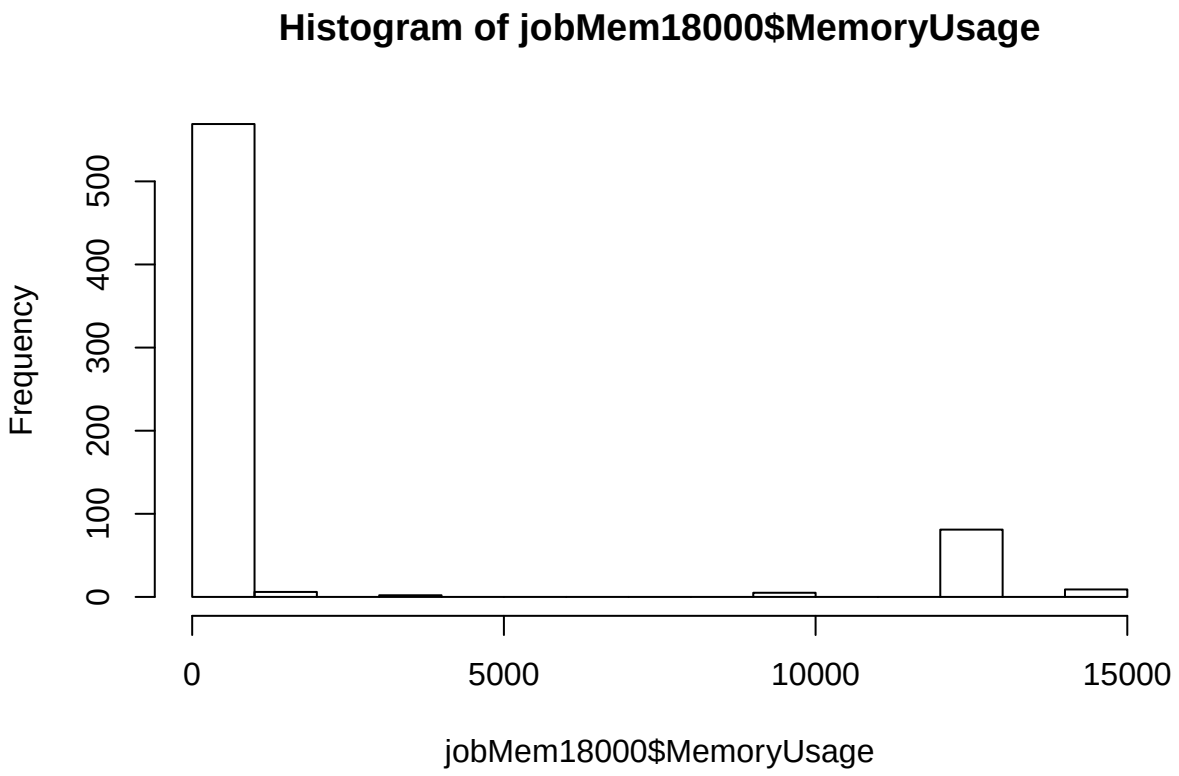
```
TotalWallTime_JobMem1900
```

```
## [1] 7578829
```

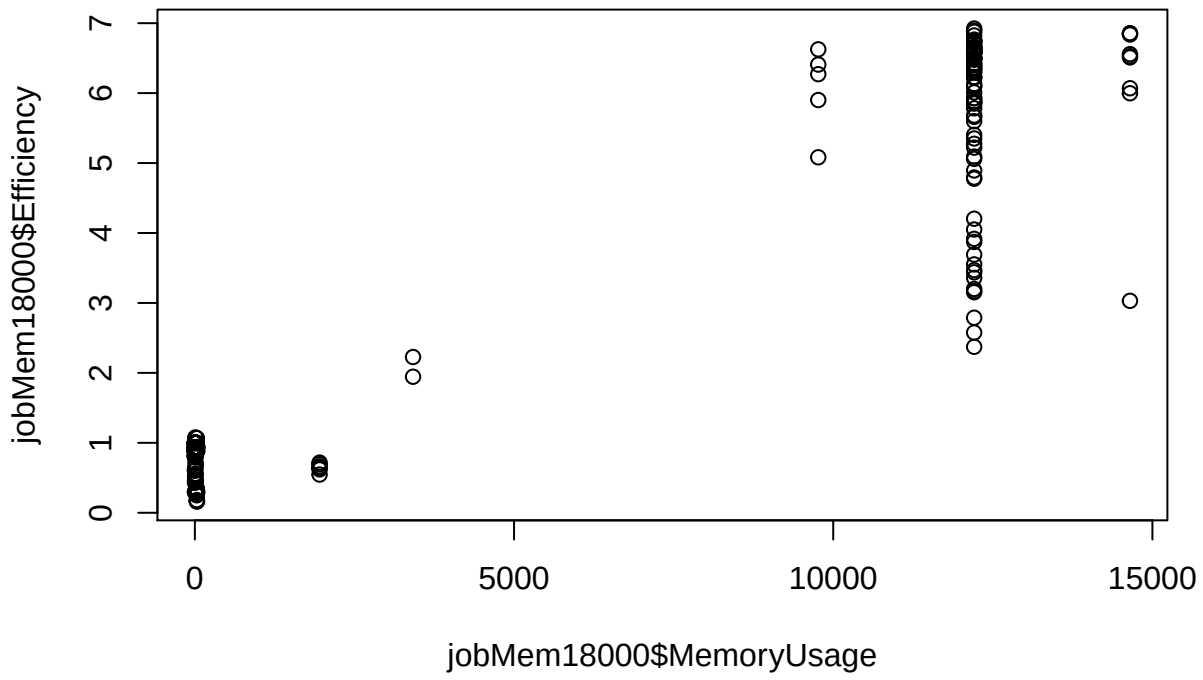
```
CumulativeEfficiency_JobMem1900 <- TotalCPUTime_JobMem1900/TotalWallTime_JobMem1900
CumulativeEfficiency_JobMem1900
```

```
## [1] 0.03041974
```

```
#################### Jobs with memory request 18000 ########################################
jobMem18000 <- subset(newdata2, newdata2$RequestMemory == 18000)
hist(jobMem18000$MemoryUsage)
```




```
plot(jobMem18000$MemoryUsage, jobMem18000$Efficiency)
```




```
TotalCPUTime_JobMem18000 <- sum(as.numeric(jobMem18000$CPUTime))
TotalWallTime_JobMem18000 <- sum(jobMem18000$WallTime)
TotalCPUTime_JobMem18000
```

```
## [1] 4608752
```

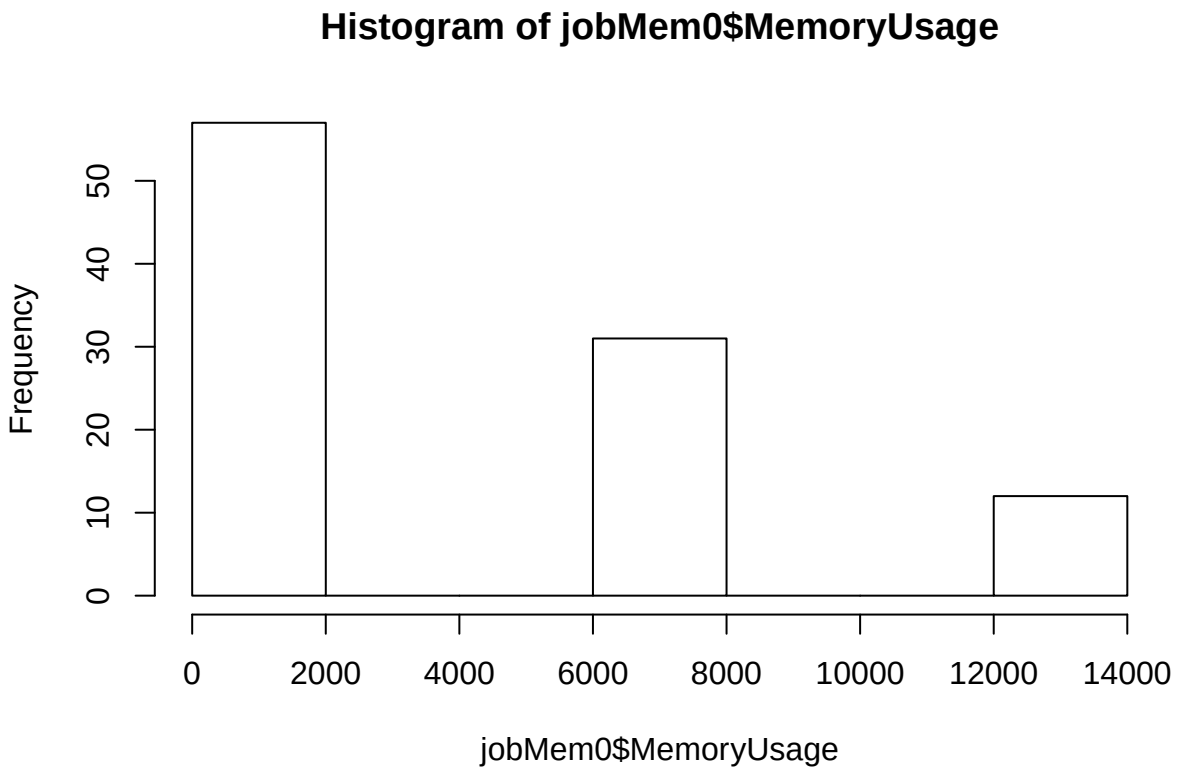
```
TotalWallTime_JobMem18000
```

```
## [1] 801418
```

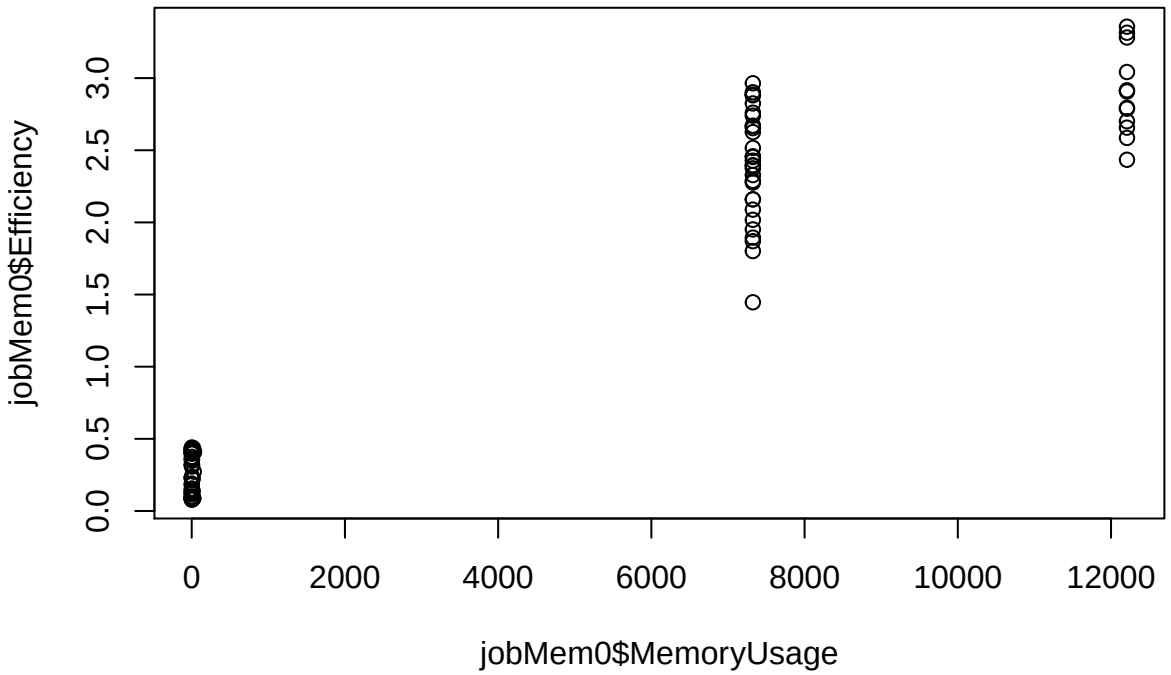
```
CumulativeEfficiency_JobMem18000 <- TotalCPUTime_JobMem18000/TotalWallTime_JobMem18000
CumulativeEfficiency_JobMem18000
```

```
## [1] 5.750747
```

```
#################### Jobs with memory request 0 ########################################
jobMem0 <- subset(newdata2, newdata2$RequestMemory == 0)
hist(jobMem0$MemoryUsage)
```




```
plot(jobMem0$MemoryUsage, jobMem0$Efficiency)
```




```
TotalCPUTime_JobMem0 <- sum(as.numeric(jobMem0$CPUTime))
TotalWallTime_JobMem0 <- sum(jobMem0$WallTime)
TotalCPUTime_JobMem0
```

```
## [1] 2155444
```

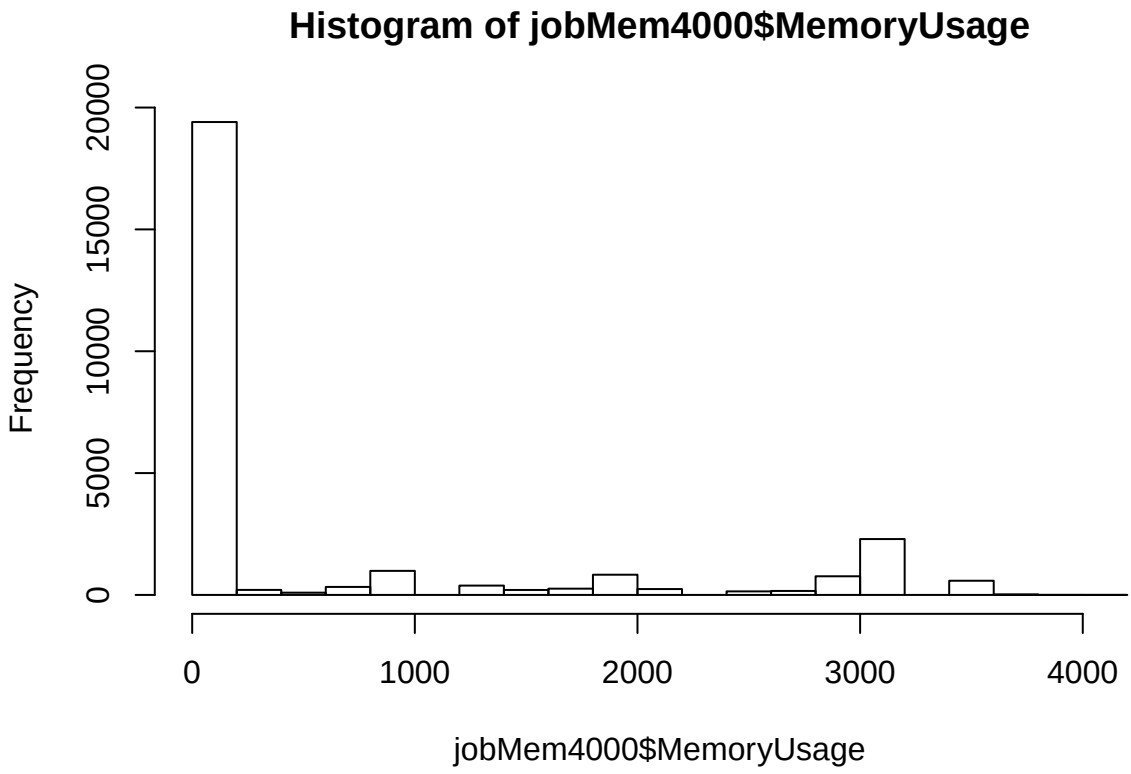
```
TotalWallTime_JobMem0
```

```
## [1] 808976
```

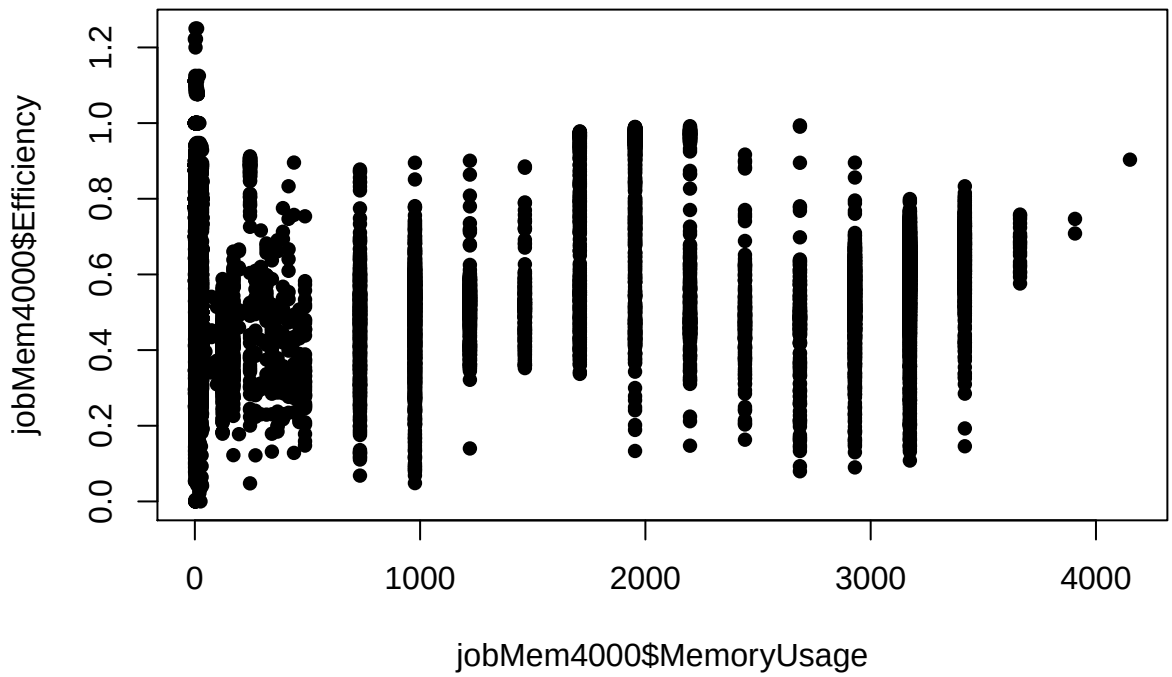
```
CumulativeEfficiency_JobMem0 <- TotalCPUTime_JobMem0/TotalWallTime_JobMem0
CumulativeEfficiency_JobMem0
```

```
## [1] 2.66441
```

```
#################### Jobs with memory request 4000 ########################################
jobMem4000 <- subset(newdata2, newdata2$RequestMemory == 4000)
hist(jobMem4000$MemoryUsage)
```




```
plot(jobMem4000$MemoryUsage, jobMem4000$Efficiency, pch=16)
```




```
TotalCPUTime_JobMem4000 <- sum(as.numeric(jobMem4000$CPUTime))
TotalWallTime_JobMem4000 <- sum(jobMem4000$WallTime)
TotalCPUTime_JobMem4000
```

```
## [1] 23008489
```

```
TotalWallTime_JobMem4000
```

```
## [1] 31687731
```

```
CumulativeEfficiency_JobMem4000 <- TotalCPUTime_JobMem4000/TotalWallTime_JobMem4000
CumulativeEfficiency_JobMem4000
```

```
## [1] 0.7261009
```

```
#################### Jobs with memory request 2000 ########################################
jobMem2000 <- subset(newdata2, newdata2$RequestMemory == 2000)
hist(jobMem2000$MemoryUsage)
```


```
plot(jobMem2000$MemoryUsage, jobMem2000$Efficiency, pch=16)
```


```
TotalCPUTime_JobMem2000 <- sum(as.numeric(jobMem2000$CPUTime))
TotalWallTime_JobMem2000 <- sum(jobMem2000$WallTime)
TotalCPUTime_JobMem2000
```

```
## [1] 4460763
```

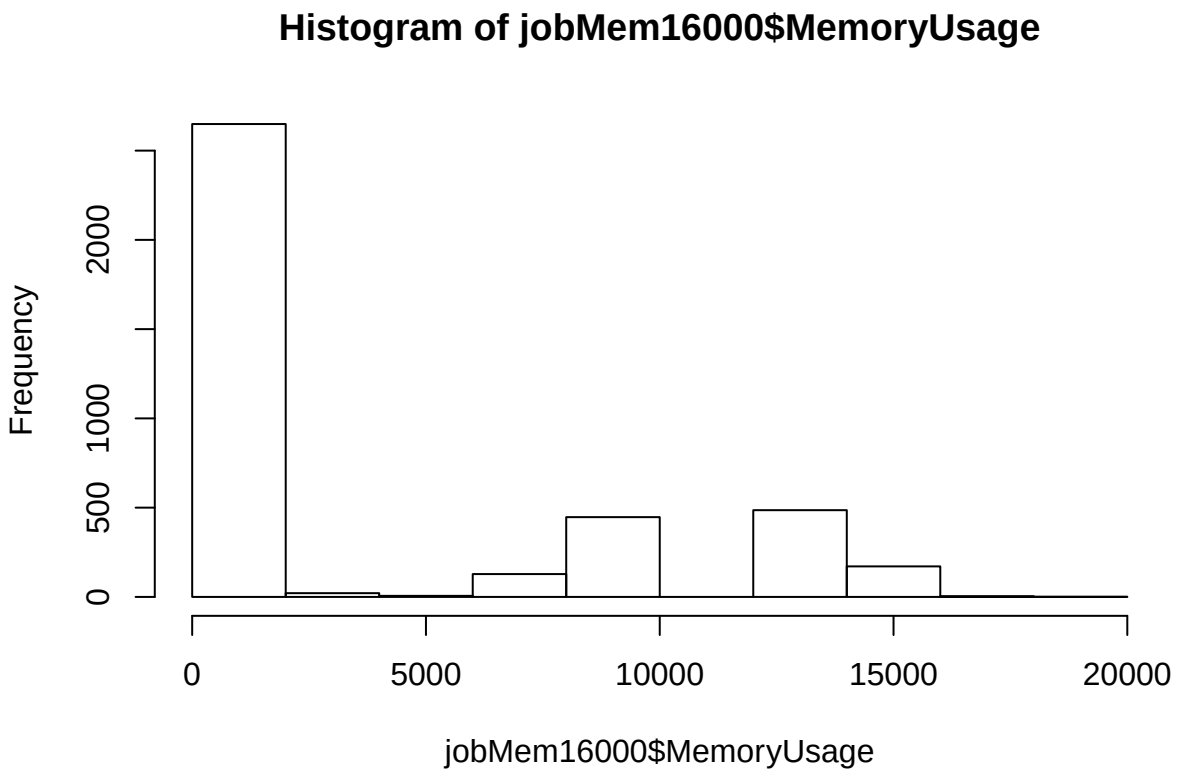
```
TotalWallTime_JobMem2000
```

```
## [1] 5868303
```

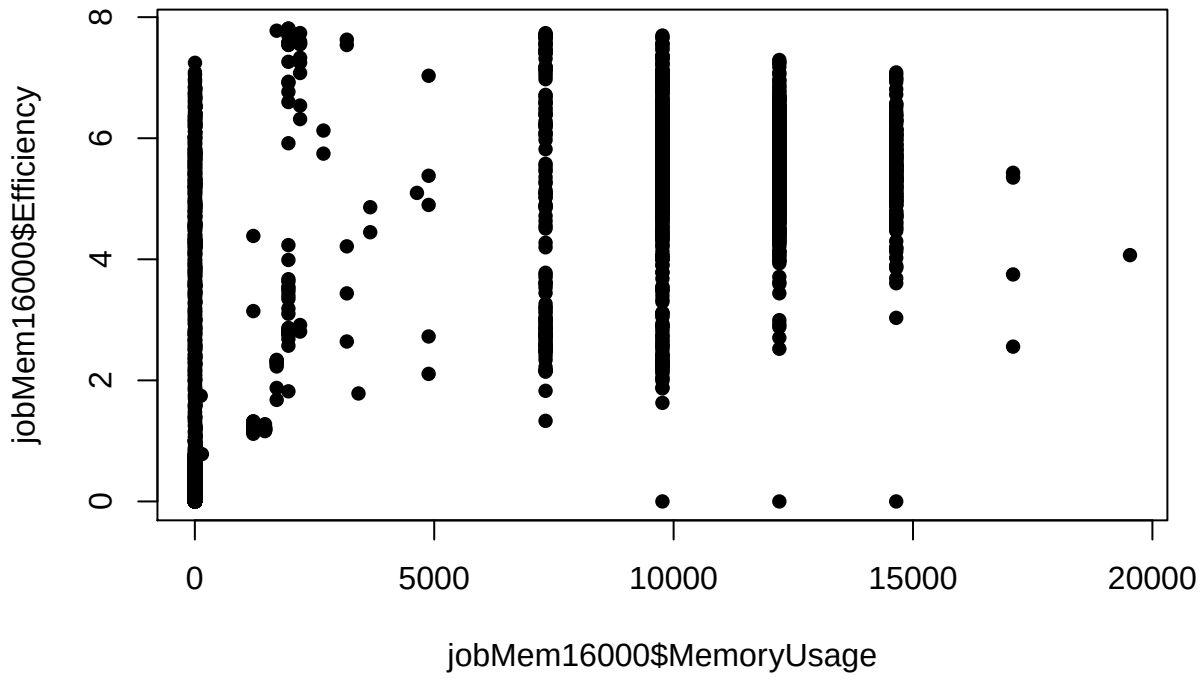
```
CumulativeEfficiency_JobMem2000 <- TotalCPUTime_JobMem2000/TotalWallTime_JobMem2000
CumulativeEfficiency_JobMem2000
```

```
## [1] 0.7601453
```

```
#################### Jobs with memory request 16000 ########################################
jobMem16000 <- subset(newdata2, newdata2$RequestMemory == 16000)
hist(jobMem16000$MemoryUsage)
```




```
plot(jobMem16000$MemoryUsage, jobMem16000$Efficiency, pch=16)
```




```
TotalCPUTime_JobMem16000 <- sum(as.numeric(jobMem16000$CPUTime))
TotalWallTime_JobMem16000 <- sum(jobMem16000$WallTime)
TotalCPUTime_JobMem16000
```

```
## [1] 1619059498
```

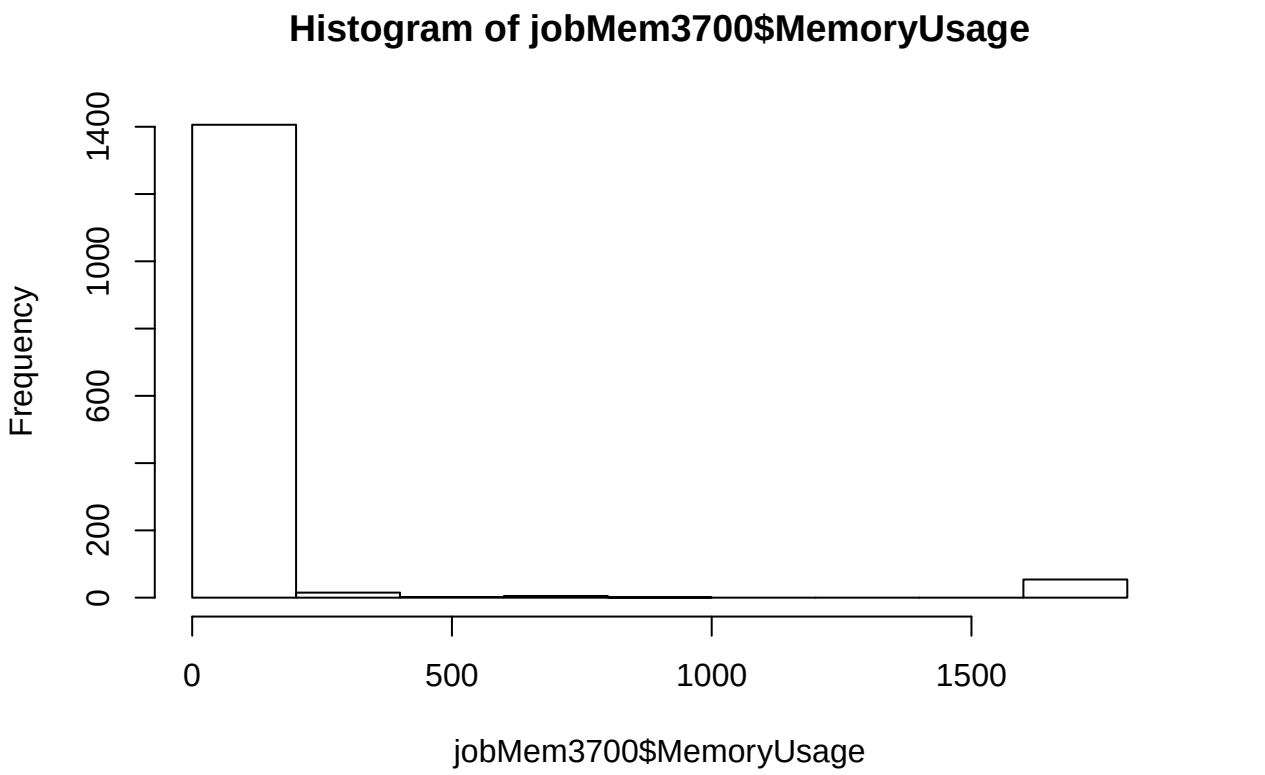
```
TotalWallTime_JobMem16000
```

```
## [1] 309173393
```

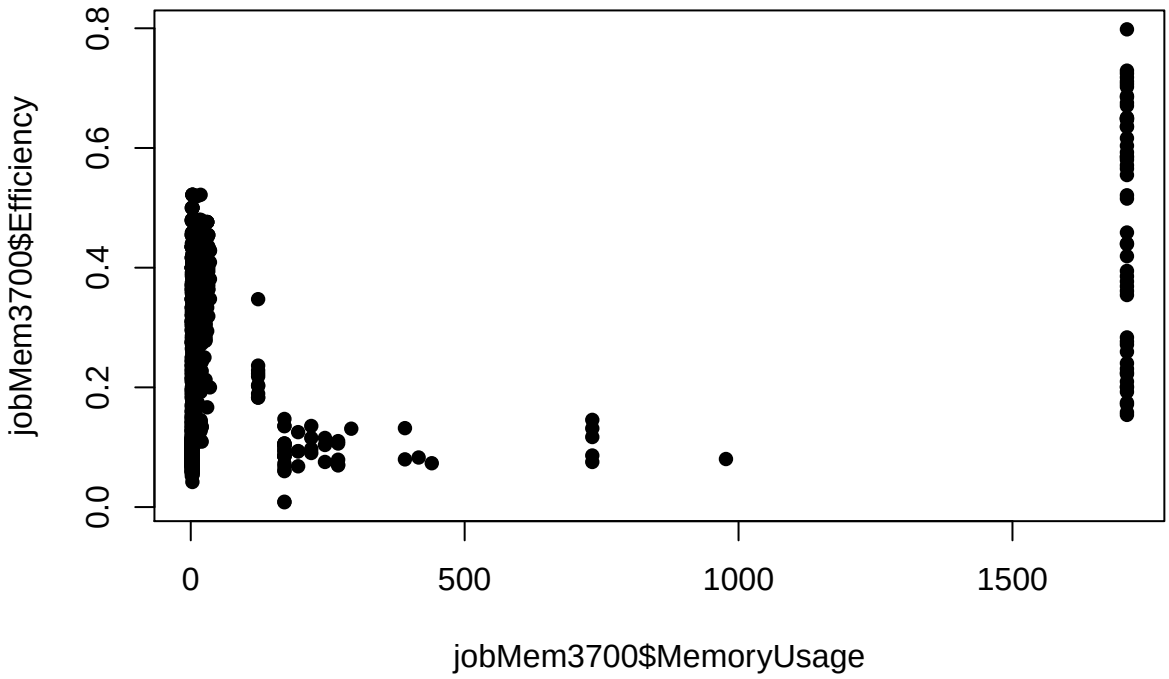
```
CumulativeEfficiency_JobMem16000 <- TotalCPUTime_JobMem16000/TotalWallTime_JobMem16000
CumulativeEfficiency_JobMem16000
```

```
## [1] 5.236736
```

```
#################### Jobs with memory request 3700 ########################################
jobMem3700 <- subset(newdata2, newdata2$RequestMemory == 3700)
hist(jobMem3700$MemoryUsage)
```




```
plot(jobMem3700$MemoryUsage, jobMem3700$Efficiency, pch=16)
```




```
TotalCPUTime_JobMem3700 <- sum(as.numeric(jobMem3700$CPUTime))
TotalWallTime_JobMem3700 <- sum(jobMem3700$WallTime)
TotalCPUTime_JobMem3700
```

```
## [1] 58722
```

```
TotalWallTime_JobMem3700
```

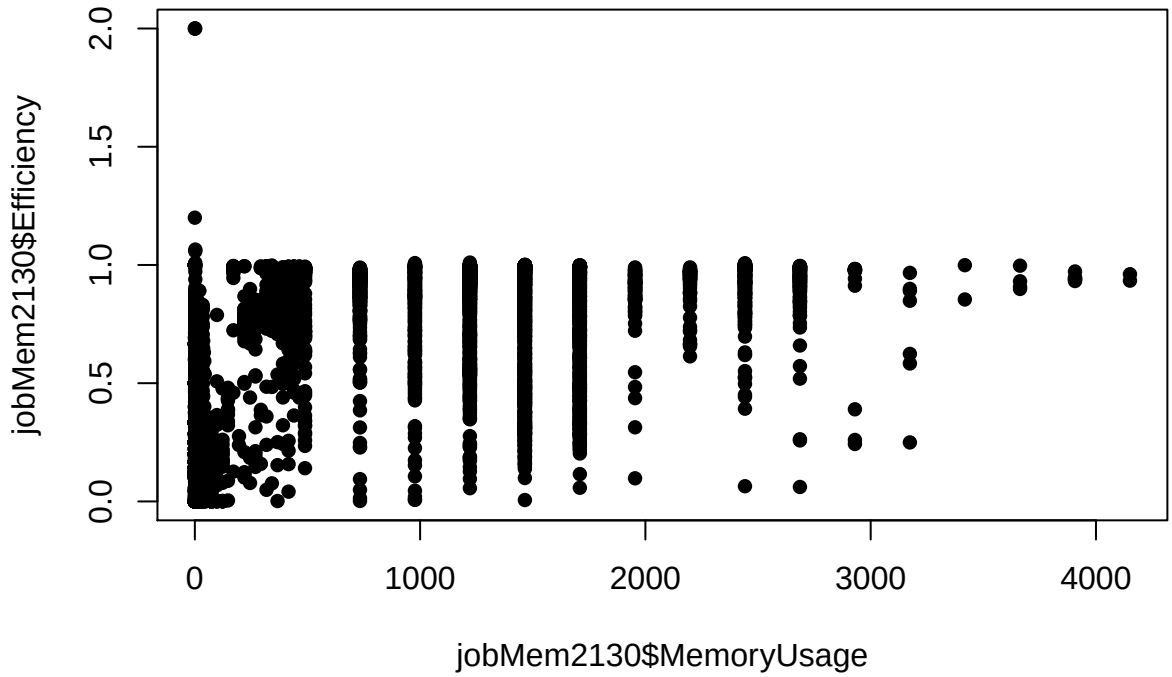
```
## [1] 248023
```

```
CumulativeEfficiency_JobMem3700 <- TotalCPUTime_JobMem3700/TotalWallTime_JobMem3700
CumulativeEfficiency_JobMem3700
```

```
## [1] 0.2367603
```

```
#################### Jobs with memory request 2130 ########################################
jobMem2130 <- subset(newdata2, newdata2$RequestMemory == 2130)
hist(jobMem2130$MemoryUsage)
```


```
plot(jobMem2130$MemoryUsage, jobMem2130$Efficiency, pch=16)
```




```
TotalCPUTime_JobMem2130 <- sum(as.numeric(jobMem2130$CPUTime))
TotalWallTime_JobMem2130 <- sum(jobMem2130$WallTime)
TotalCPUTime_JobMem2130
```

```
## [1] 243862330
```

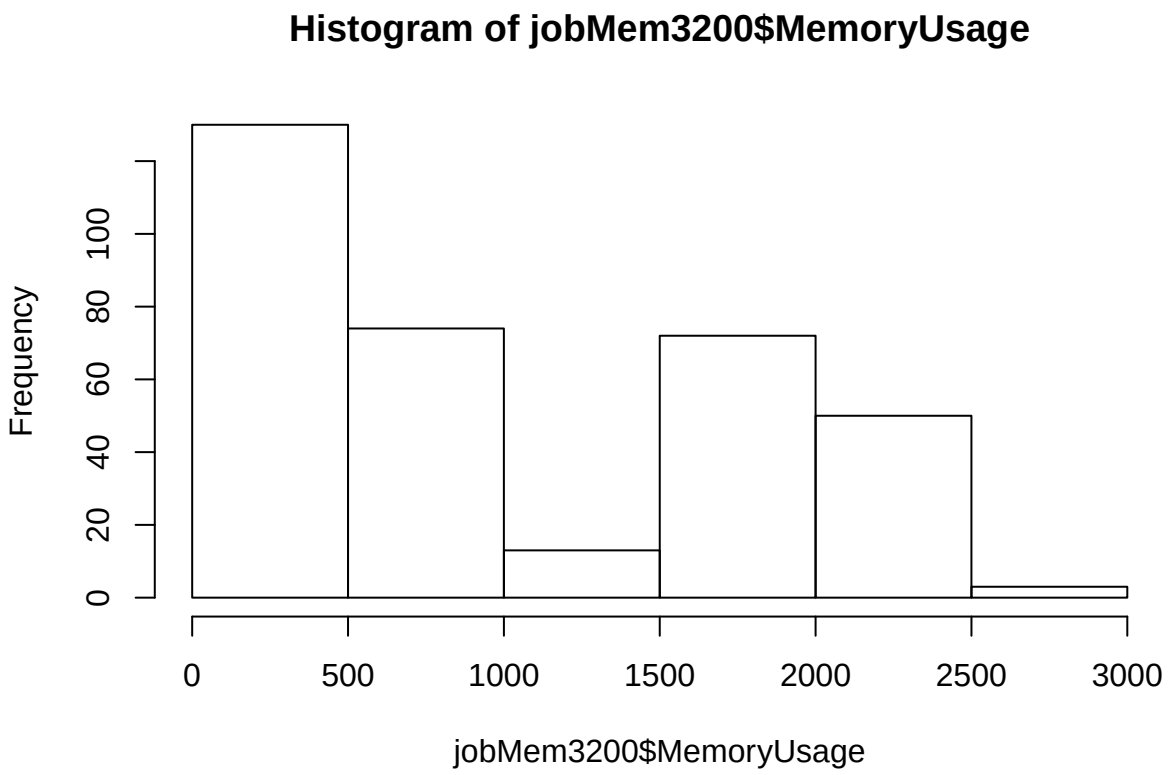
```
TotalWallTime_JobMem2130
```

```
## [1] 265954233
```

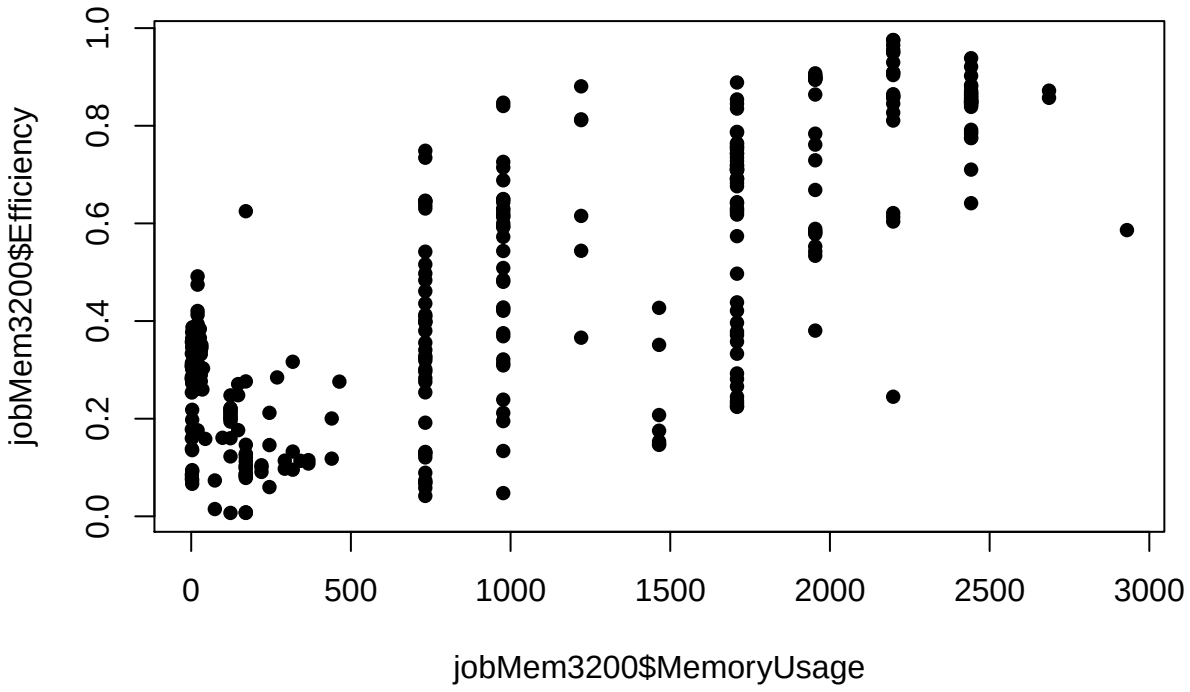
```
CumulativeEfficiency_JobMem2130 <- TotalCPUTime_JobMem2130/TotalWallTime_JobMem2130
CumulativeEfficiency_JobMem2130
```

```
## [1] 0.9169334
```

```
#################### Jobs with memory request 3200 ########################################
jobMem3200 <- subset(newdata2, newdata2$RequestMemory == 3200)
hist(jobMem3200$MemoryUsage)
```




```
plot(jobMem3200$MemoryUsage, jobMem3200$Efficiency, pch=16)
```




```
TotalCPUTime_JobMem3200 <- sum(as.numeric(jobMem3200$CPUTime))
TotalWallTime_JobMem3200 <- sum(jobMem3200$WallTime)
TotalCPUTime_JobMem3200
```

```
## [1] 1784724
```

```
TotalWallTime_JobMem3200
```

```
## [1] 2399924
```

```
CumulativeEfficiency_JobMem3200 <- TotalCPUTime_JobMem3200/TotalWallTime_JobMem3200
CumulativeEfficiency_JobMem3200
```

```
## [1] 0.7436585
```