EfficiencyJobUniverse.R

arcs

Fri Nov 24 17:49:54 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
```

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

unique(newdata$JobUniverse)
```

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

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

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

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

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

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

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

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

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

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

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

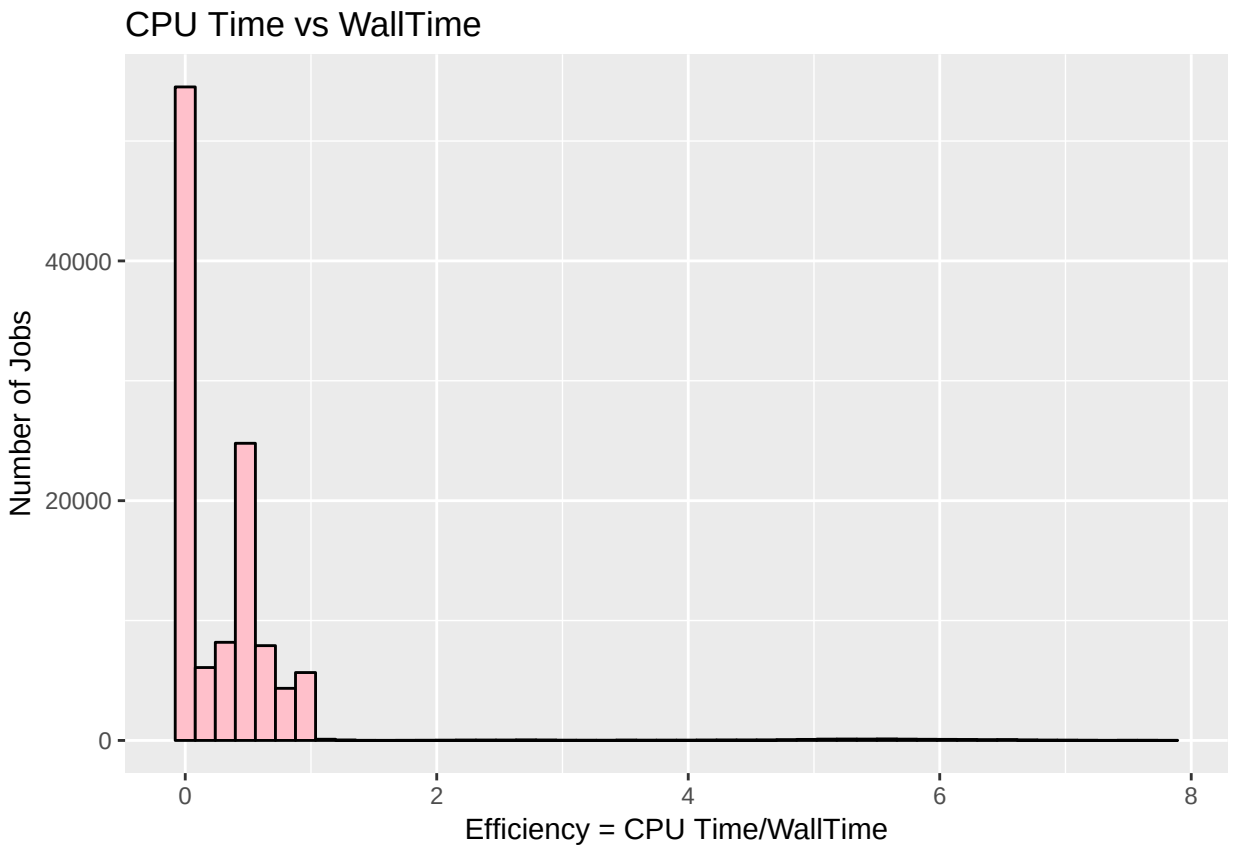
```r
########################################################################
############### Computation of efficiency ##############################
########################################################################

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:115272   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:115272   3rd Qu.:0
##  Max.   :301559                     Max.   :0                 Max.   :0
##  NA's   :1569
##   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   :   469.2   Mean   :  16007   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   27     :45698   Min.   :2130      1900   :54248
##  1st Qu.: 1900   3      : 7048   1st Qu.:2130      4000   :26909
##  Median : 2000   0      : 6476   Median :2130      2000   :13392
##  Mean   : 3044   2      : 6297   Mean   :2130      None   :11760
##  3rd Qu.: 4000   13     : 5387   3rd Qu.:2130      16000  : 3913
```

```
##  Max.   :18000   8      : 4733   Max.   :2130      0      : 1669
##                 (Other):39633                     (Other): 3381
##  CumulativeRemoteSysCpu CumulativeRemoteUserCpu Remote_JobUniverse
##  None   :35103          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):51162
##   JobUniverse    CPUTime            WallTime        Efficiency
##  Min.   :5    Min.   :      0   Min.   :     0   Min.   :0.0000
##  1st Qu.:5    1st Qu.:      4   1st Qu.:    31   1st Qu.:0.0286
##  Median :5    Median :      7   Median :   140   Median :0.1424
##  Mean   :5    Mean   :  16476   Mean   :  5493   Mean   :   Inf
##  3rd Qu.:5    3rd Qu.:     16   3rd Qu.:   144   3rd Qu.:0.5000
##  Max.   :5    Max.   :1941994   Max.   :301559   Max.   :   Inf
##                                 NA's   :1569     NA's   :1666
```

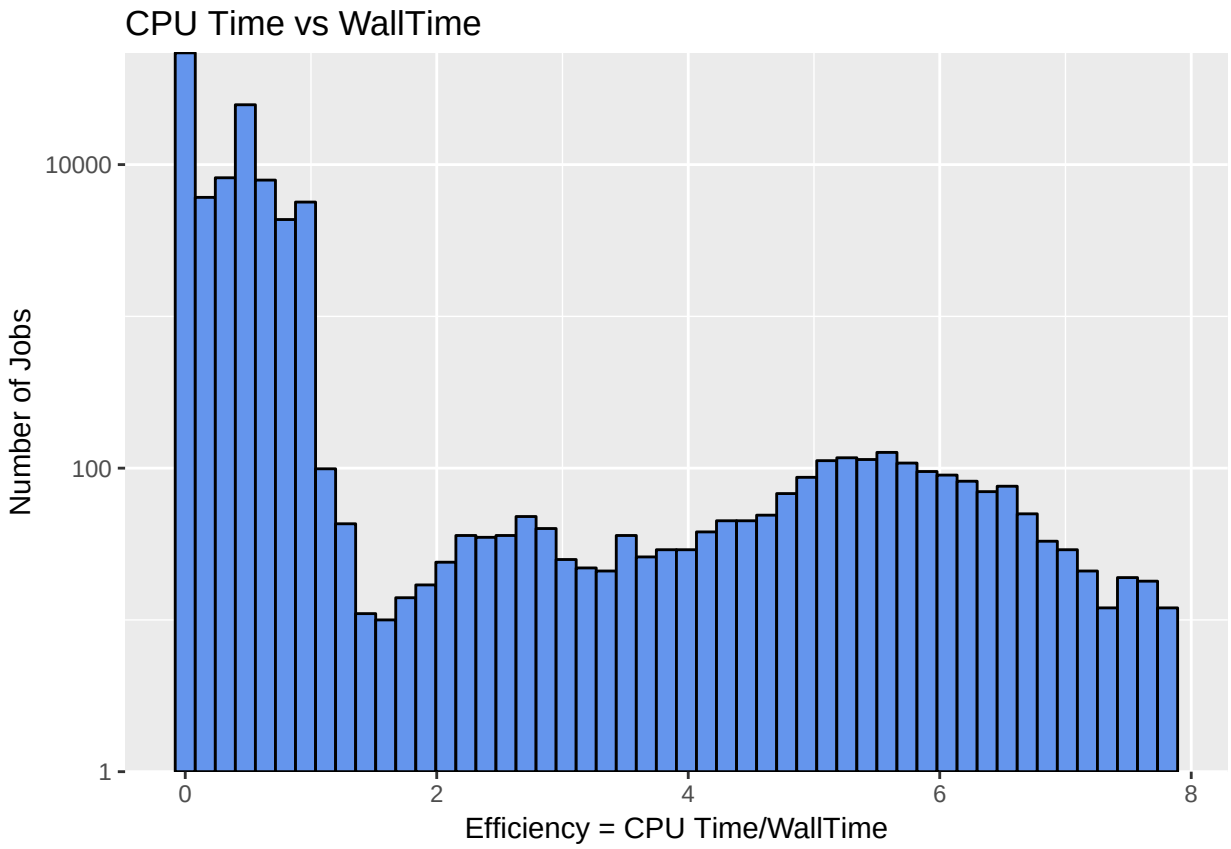
```
#Graph of CPU Time VS WallTime

graph1 <- ggplot(newdata3, aes(x = Efficiency)) +
  geom_histogram( color = "Black", fill = "Pink", bins = 50 ) #+
#scale_x_continuous(bandwidth = 0.1 )
graph1 + labs(title= "CPU Time vs WallTime", x= "Efficiency = CPU Time/WallTime", y = "Number of Jobs")
```

```
## Warning: Removed 3 rows containing non-finite values (stat_bin).
```



```
#Graph of CPU Time vs Wall Time where y=log10(x)

graph2 <- ggplot(newdata3, aes(x = Efficiency)) +
  geom_histogram(color = "Black", fill = "cornflowerblue", bins = 50 ) +
  scale_y_continuous(trans="log10", expand=c(0,0)) #+
#scale_x_continuous( xlim = c(0,1.2), bandwidth = 0.1)
graph2 + labs(title= "CPU Time vs WallTime", x= "Efficiency = CPU Time/WallTime", y = "Number of Jobs")
```

```
## Warning: Removed 3 rows containing non-finite values (stat_bin).
```




```
##########################################################################
###################### Classification of dataset ########################
#####################   into 2 groups based of  ##########################
#####################     efficiency            ##########################
##########################################################################

efficiency_grt_1.2 <- subset(newdata3, Efficiency > 1.2) #Extract Jobs with efficiency > 1.2
efficiency_less_1.2 <- subset(newdata3, Efficiency <= 1.2) #Extract Jobs with efficiency <= 1.2

##########################################################################
################ Compute Overall efficiency of    ####################
#####################   of different classes    ##########################
##########################################################################

TotalCPUTime_less_1.2 <- sum(as.numeric(efficiency_less_1.2$CPUTime))
TotalWallTime_less_1.2 <- sum(efficiency_less_1.2$WallTime)
TotalCPUTime_less_1.2
```

```
## [1] 278473597
```

```
TotalWallTime_less_1.2
```

```
## [1] 318881655
```

```
CumulativeEfficiency_less_1.2 <- TotalCPUTime_less_1.2/TotalWallTime_less_1.2
CumulativeEfficiency_less_1.2
```

```
## [1] 0.873282
```

```
TotalCPUTime_grt_1.2 <- sum(as.numeric(efficiency_grt_1.2$CPUTime))
TotalWallTime_grt_1.2 <- sum(efficiency_grt_1.2$WallTime)
TotalCPUTime_grt_1.2
```

```
## [1] 1620764469
```

```
TotalWallTime_grt_1.2
```

```
## [1] 305659237
```

```
CumulativeEfficiency_grt_1.2 <- TotalCPUTime_grt_1.2/TotalWallTime_grt_1.2
CumulativeEfficiency_grt_1.2
```

```
## [1] 5.302521
```

```
TotalCPUTime <- sum(as.numeric(newdata3$CPUTime))
TotalWallTime <- sum(newdata3$WallTime)
TotalCPUTime
```

```
## [1] 1899238066
```

```
TotalWallTime
```

```
## [1] 624540892
```

```
CumulativeEfficiency <- TotalCPUTime/TotalWallTime
CumulativeEfficiency
```

```
## [1] 3.041015
```

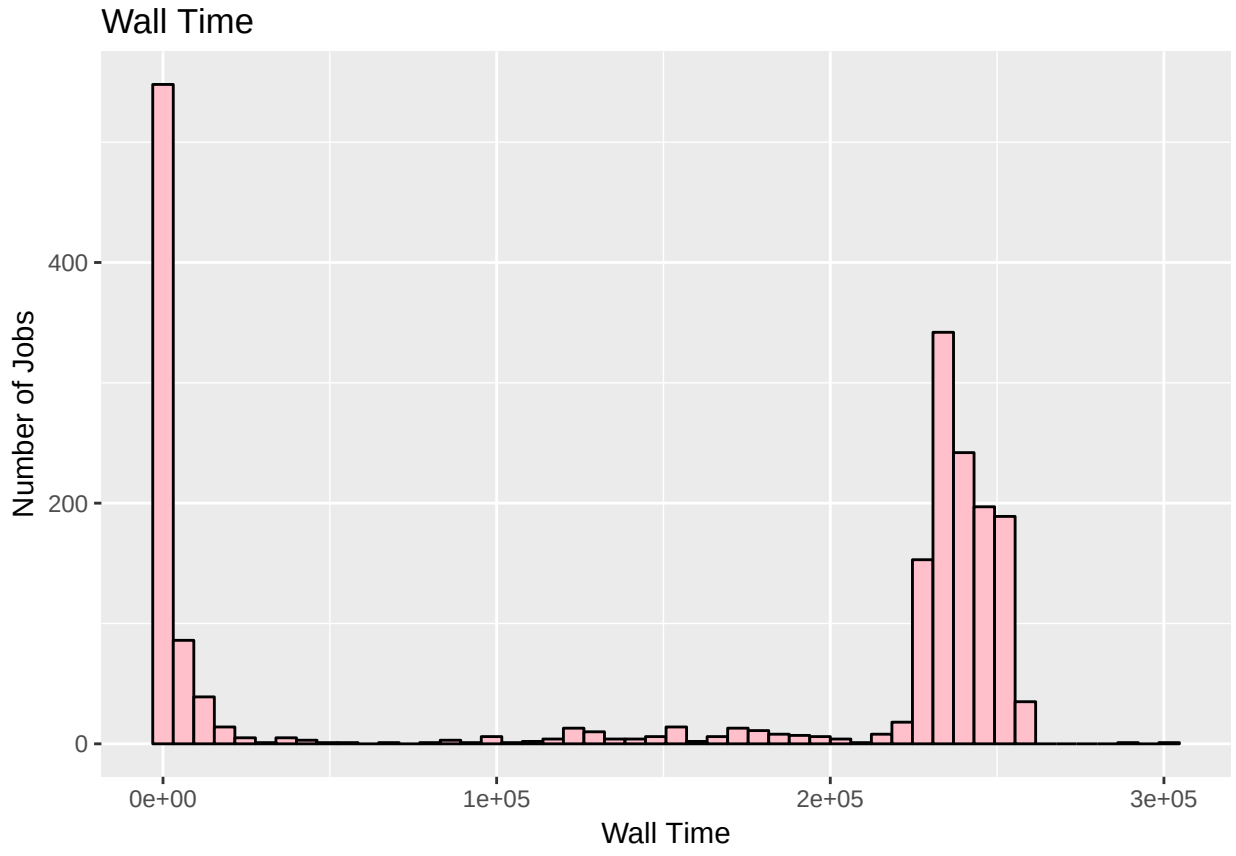
```
##########################################################################################
#####################################   SUMMARY    #######################################
##########################################################################################
############################  Total no of jobs : 257561 ##################################
########################  Number of jobs after cleansing : 113703 ########################
##############  Number of jobs with efficiency <= 1.2 (Class I) : 111589 #################
###############  Number of jobs with efficiency > 1.2 (Class II) : 2017 ##################
############################  Overall Efficiency : 3.041 #################################
#####################  Overall Efficiency of Class I jobs : 0.873 ########################
#####################  Overall Efficiency of Class I jobs : 5.302 ########################
##########################################################################################
```

```
#Graph of CPU Time for jobs with efficiency >1.2

png(filename = "/home/arcs/Oct14/RCodes/Plots/CPUTime.png")


graph3 <- ggplot(efficiency_grt_1.2, aes(x = CPUTime)) +
```

```
  geom_histogram( color = "Black", fill = "Pink", bins = 50 ) #+
#scale_x_continuous(bandwidth = 0.1 )
graph3 + labs(title= "CPU Time", x= "CPU Time", y = "Number of Jobs")
dev.off()
```

```
## pdf
##   2
```

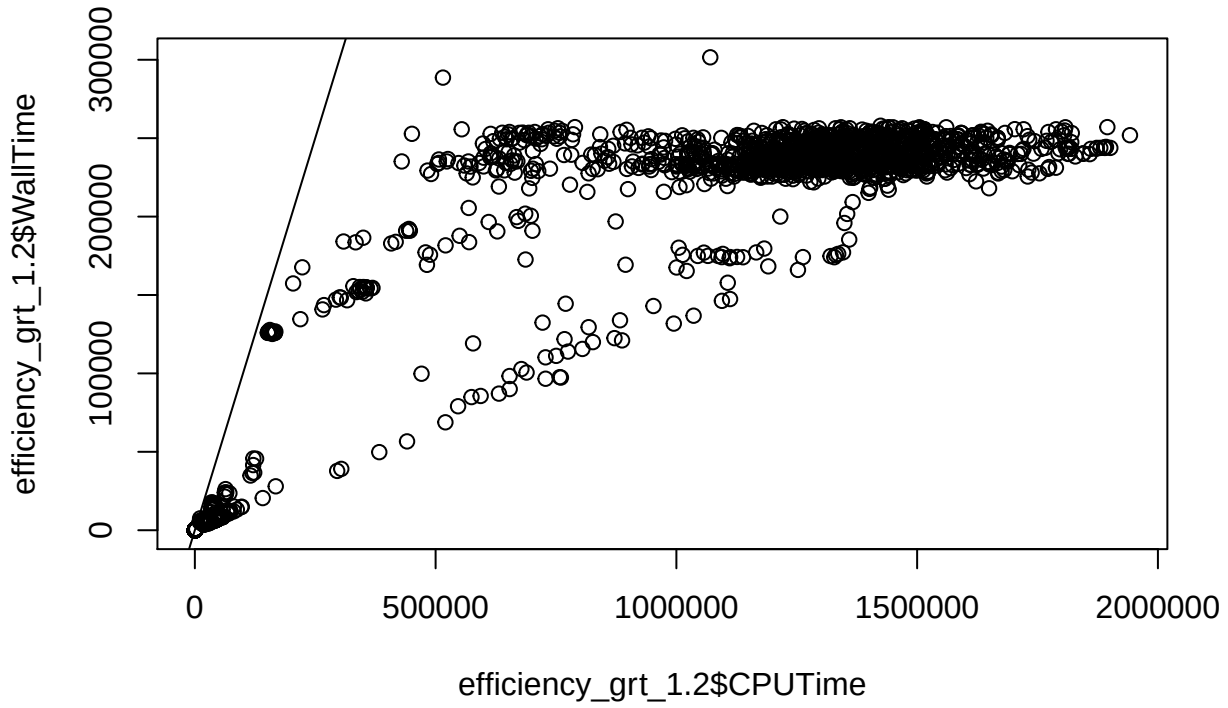
```
#Graph of Wall Time for jobs with efficiency >1.2

graph4 <- ggplot(efficiency_grt_1.2, aes(x = WallTime)) +
  geom_histogram( color = "Black", fill = "Pink", bins = 50 ) #+
#scale_x_continuous(bandwidth = 0.1 )
graph4 + labs(title= "Wall Time", x= "Wall Time", y = "Number of Jobs")
```



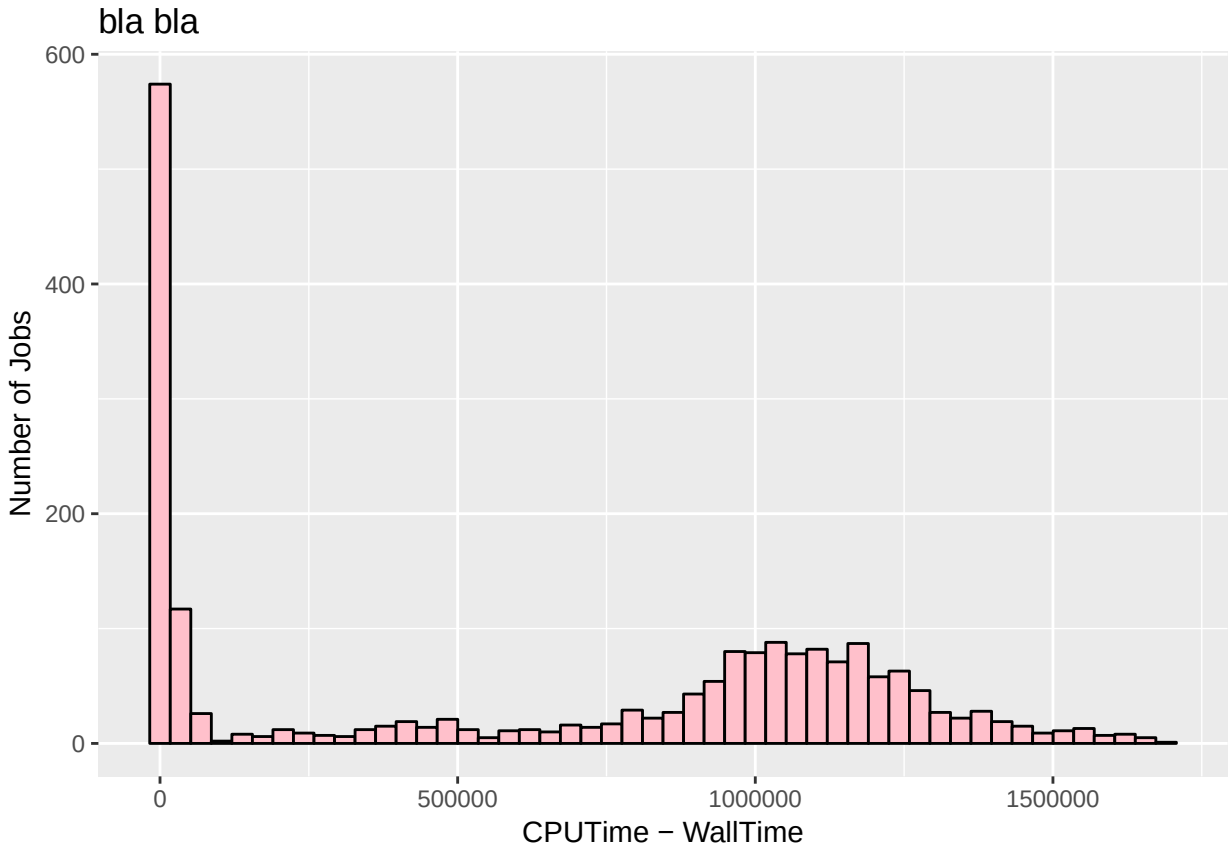

```
plot(efficiency_grt_1.2$CPUTime, efficiency_grt_1.2$WallTime, alpha = .5)
```

```
## Warning in plot.window(...): "alpha" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "alpha" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "alpha" is not
## a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "alpha" is not
## a graphical parameter

## Warning in box(...): "alpha" is not a graphical parameter

## Warning in title(...): "alpha" is not a graphical parameter
```

```
abline(a = 0, b = 1)
```




```
efficiency_grt_1.2$Diff <- efficiency_grt_1.2$CPUTime - efficiency_grt_1.2$WallTime

graph5 <- ggplot(efficiency_grt_1.2, aes(x = Diff)) +
  geom_histogram( color = "Black", fill = "Pink", bins = 50 ) #+
#scale_x_continuous(bandwidth = 0.1 )
graph5 + labs(title= "bla bla", x= "CPUTime - WallTime", y = "Number of Jobs")
```




```
Error <- sum(efficiency_grt_1.2$Diff)
```