

Machine utilization

Investigate one of their heavy machines - RL1 You have been supplied one month worth of data for all of their machines. The dataset shows what percentage of capacity for each machine was idle (unused) in any given hour. You are required to deliver an R list with the following components: Character: Machine name Vector: (min, mean, max) utilisation for the month (excluding unknown hours) Logical: Has utilisation ever fallen below 90%? TRUE / FALSE Vector: All hours where utilisation is unknown (NAs) Dataframe: For this machine Plot: For all machines

```
getwd()
```

```
## [1] "C:/Users/badal/Desktop/Advance data science/R Programming Advanced Analytics In R For Data Science/r-adv"
```

```
r11 <- read.csv("file:///C://Users//badal/Desktop//dataset_//P3-Machine-Utilization.csv")
```

```
head(r11)
```

```
##           Timestamp Machine Percent.Idle
## 1 01/09/2016 00:00      RL1           NA
## 2 01/09/2016 01:00      RL1           NA
## 3 01/09/2016 02:00      RL1           NA
## 4 01/09/2016 03:00      RL1           NA
## 5 01/09/2016 04:00      RL1           NA
## 6 01/09/2016 05:00      RL1           NA
```

```
str(r11)
```

```
## 'data.frame':   3600 obs. of  3 variables:
## $ Timestamp    : Factor w/ 720 levels "01/09/2016 00:00",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Machine      : Factor w/ 5 levels "RL1","RL2","SR1",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Percent.Idle: num  NA NA NA NA NA ...
```

```
r11
```

```
str(r11)
```

```
## 'data.frame':   3600 obs. of  3 variables:
## $ Timestamp    : Factor w/ 720 levels "01/09/2016 00:00",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Machine      : Factor w/ 5 levels "RL1","RL2","SR1",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Percent.Idle: num  NA NA NA NA NA ...
```

```
r11$utilazation <- 1-r11$Percent.Idle
```

```
head(r11,10)
```

```
##          Timestamp Machine Percent.Idle utilazation
## 1 01/09/2016 00:00      RL1          NA          NA
## 2 01/09/2016 01:00      RL1          NA          NA
## 3 01/09/2016 02:00      RL1          NA          NA
## 4 01/09/2016 03:00      RL1          NA          NA
## 5 01/09/2016 04:00      RL1          NA          NA
## 6 01/09/2016 05:00      RL1          NA          NA
## 7 01/09/2016 06:00      RL1          NA          NA
## 8 01/09/2016 07:00      RL1 0.01994048 0.9800595
## 9 01/09/2016 08:00      RL1 0.01997024 0.9800298
## 10 01/09/2016 09:00      RL1 0.02119048 0.9788095
```

Handling date-times in R

```
rl1$Timestamp <- as.POSIXct(rl1$Timestamp ,format= "%d/%m/%Y %H:%M")
summary(rl1)
```

```
##      Timestamp                Machine      Percent.Idle
## Min.   :2016-09-01 00:00:00    RL1 :720    Min.   :0.0000
## 1st Qu.:2016-09-08 11:45:00    RL2 :720    1st Qu.:0.0262
## Median :2016-09-15 23:30:00    SR1 :720    Median :0.0410
## Mean   :2016-09-15 23:30:00    SR4A:720    Mean   :0.0431
## 3rd Qu.:2016-09-23 11:15:00    SR6 :720    3rd Qu.:0.0576
## Max.   :2016-09-30 23:00:00                Max.   :0.1508
##                                     NA's   :361
##      utilazation
## Min.   :0.8492
## 1st Qu.:0.9424
## Median :0.9590
## Mean   :0.9569
## 3rd Qu.:0.9738
## Max.   :1.0000
## NA's   :361
```

```
RL1 <- rl1[rl1$Machine=="RL1",]
RL1$Machine <- factor(RL1$Machine)
summary(RL1)
```

```
##      Timestamp                Machine      Percent.Idle
## Min.   :2016-09-01 00:00:00    RL1:720    Min.   :0.00500
## 1st Qu.:2016-09-08 11:45:00                1st Qu.:0.03208
## Median :2016-09-15 23:30:00                Median :0.04613
## Mean   :2016-09-15 23:30:00                Mean   :0.04830
## 3rd Qu.:2016-09-23 11:15:00                3rd Qu.:0.05967
## Max.   :2016-09-30 23:00:00                Max.   :0.15077
##                                     NA's   :7
##      utilazation
## Min.   :0.8492
## 1st Qu.:0.9403
## Median :0.9539
## Mean   :0.9517
```

```
## 3rd Qu.:0.9679
## Max. :0.9950
## NA's :7

RL1_stat_utl <- c(min(RL1$utilazation, na.rm = T),
                  mean(RL1$utilazation, na.rm = T),
                  max(RL1$utilazation, na.rm = T))

RL1_stat_utl

## [1] 0.8492262 0.9516976 0.9950000
```

logical : Has utilization ever fallen below 90% ? T/F

```
which(RL1$utilazation < 0.90)

## [1] 31 32 33 34 35 36 37 326 327 328 329 330 331 332 484 485 486
## [18] 487 488 489 490 491 702 703 704 705 706
```

how many time utilization gone under 90%

```
util_under_90_flag <- length(which(RL1$utilazation < 0.90)) > 0
util_under_90_flag

## [1] TRUE
```

working with list

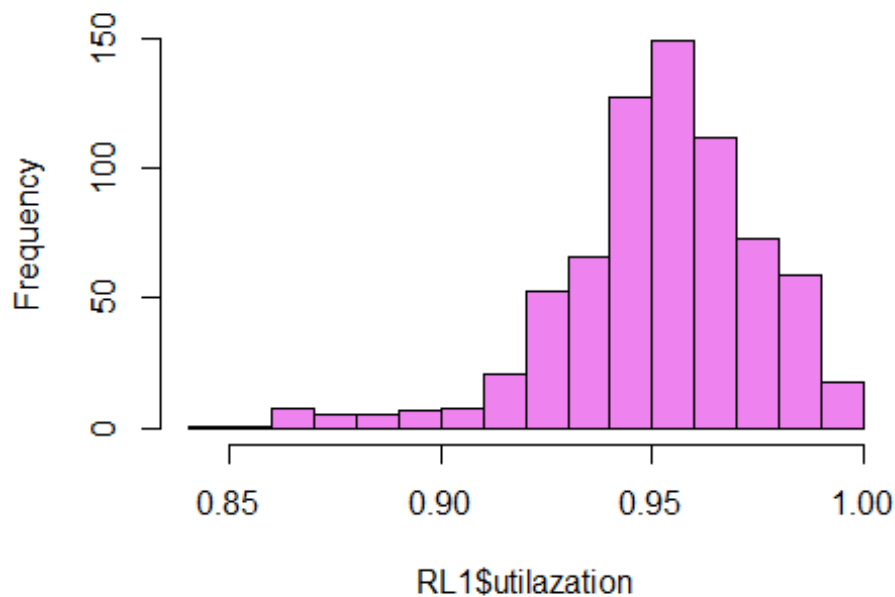
```
list_RL1 <- list(Machine = "RL1", Stats = RL1_stat_utl,
                 LowThrashold = util_under_90_flag )

list_RL1

## $Machine
## [1] "RL1"
##
## $Stats
## [1] 0.8492262 0.9516976 0.9950000
##
## $LowThrashold
## [1] TRUE

hist(RL1$utilazation, col = "violet")
```

Histogram of RL1\$utilazation



```
list_RL1$unknownhrs <- RL1[is.na(RL1$utilazation),"Timestamp"]
list_RL1

## $Machine
## [1] "RL1"
##
## $Stats
## [1] 0.8492262 0.9516976 0.9950000
##
## $LowThrashold
## [1] TRUE
##
## $unknownhrs
## [1] "2016-09-01 00:00:00 IST" "2016-09-01 01:00:00 IST"
## [3] "2016-09-01 02:00:00 IST" "2016-09-01 03:00:00 IST"
## [5] "2016-09-01 04:00:00 IST" "2016-09-01 05:00:00 IST"
## [7] "2016-09-01 06:00:00 IST"

list_RL1[[4]][3]

## [1] "2016-09-01 02:00:00 IST"

list_RL1$data <- RL1
summary(list_RL1)

##           Length Class      Mode
## Machine      1    -none- character
## Stats         3    -none-  numeric
```

```
## LowThrashold 1      -none-      logical
## unknownhrs    7      POSIXct     numeric
## data          4      data.frame  list
```

TIME-STAMP

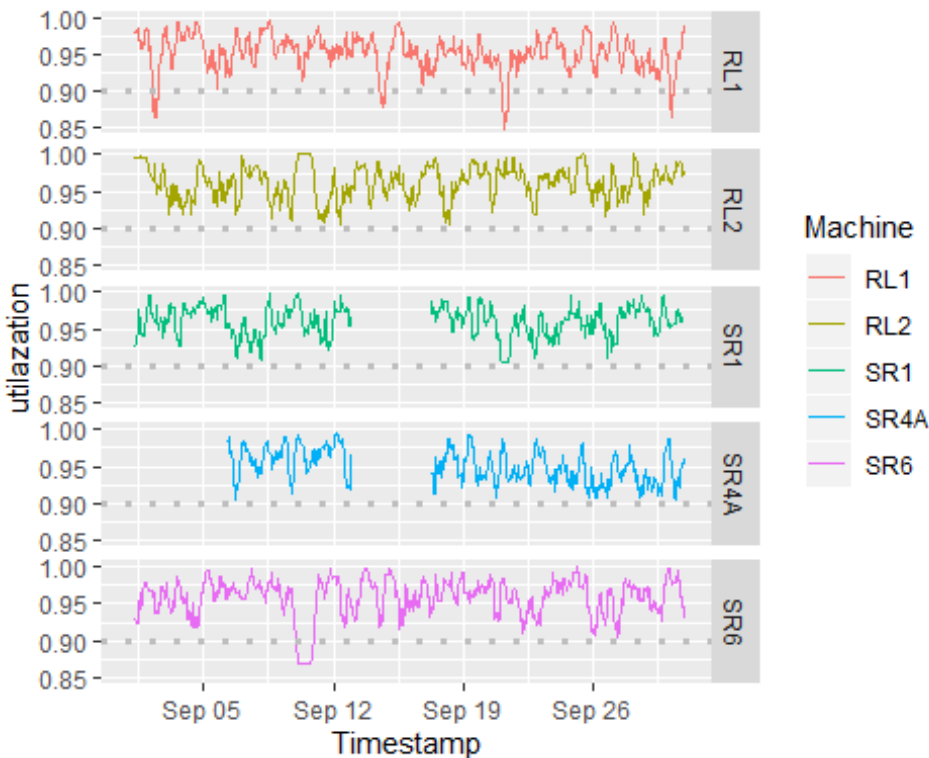
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```

```
plot <- ggplot(r11)
myplot <- plot + geom_line(aes(x = Timestamp, y = utilization, color =
Machine), size = 0.5) + facet_grid(Machine~.) +
  geom_hline(yintercept = 0.90, color='Grey', size=1.2,
            linetype =3)
```

```
myplot
```

```
## Warning: Removed 155 rows containing missing values (geom_path).
```



```
list_RL1$Plot <- myplot
```

```
list_RL1
```

```
## $Machine
```

```
## [1] "RL1"
```

```
##
```

```
## $Stats
```

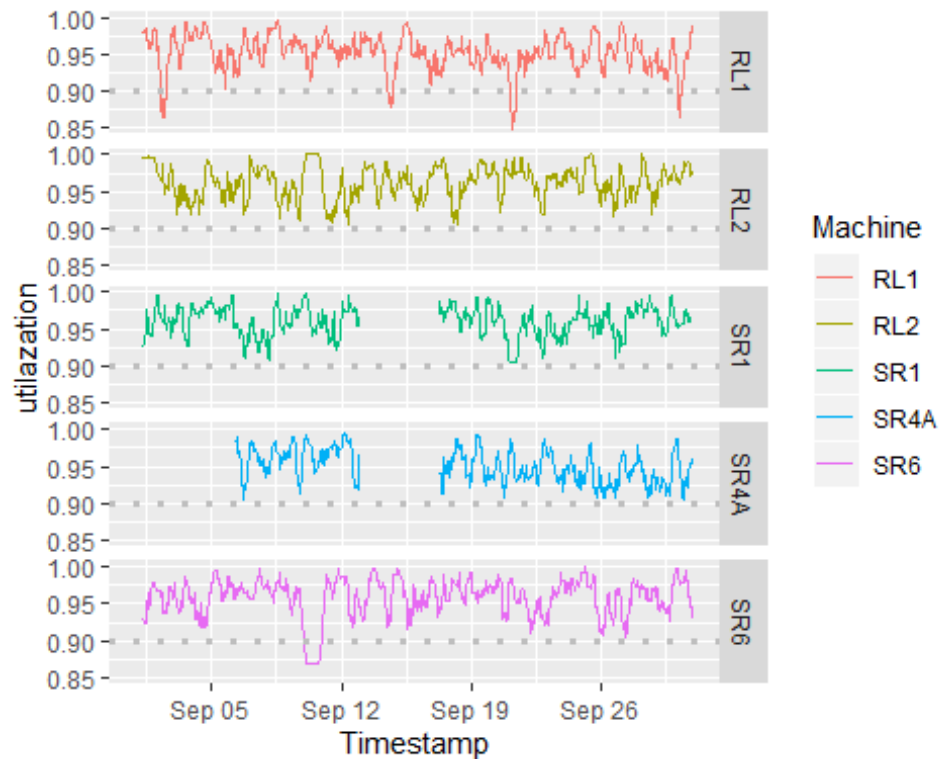
```
## [1] 0.8492262 0.9516976 0.9950000
```

```
##
```

```
## $LowThrashold
## [1] TRUE
##
## $unknownhrs
## [1] "2016-09-01 00:00:00 IST" "2016-09-01 01:00:00 IST"
## [3] "2016-09-01 02:00:00 IST" "2016-09-01 03:00:00 IST"
## [5] "2016-09-01 04:00:00 IST" "2016-09-01 05:00:00 IST"
## [7] "2016-09-01 06:00:00 IST"
##
## $data
##           Timestamp Machine Percent.Idle utilazation
## 1    2016-09-01 00:00:00      RL1          NA         NA
## 2    2016-09-01 01:00:00      RL1          NA         NA
## 3    2016-09-01 02:00:00      RL1          NA         NA
## 4    2016-09-01 03:00:00      RL1          NA         NA
## 5    2016-09-01 04:00:00      RL1          NA         NA
## 6    2016-09-01 05:00:00      RL1          NA         NA
## 7    2016-09-01 06:00:00      RL1          NA         NA
## 8    2016-09-01 07:00:00      RL1   0.019940476   0.9800595
## 9    2016-09-01 08:00:00      RL1   0.019970238   0.9800298
## 10   2016-09-01 09:00:00      RL1   0.021190476   0.9788095
.
```

```
## $Plot
```

```
## Warning: Removed 155 rows containing missing values (geom_path).
```



```
summary(list_RL1)
```

```
##           Length Class      Mode
## Machine      1    -none-  character
## Stats        3    -none-  numeric
## LowThreshold 1    -none-  logical
## unknownhrs   7   POSIXct  numeric
## data         4   data.frame list
## Plot         9     gg      list
```