exercise1.R

jun

Thu Jul 7 19:09:10 2016

```
##practice

##vector
x <- c(1,2)                 #integer
x <- c(1.7,2.4)             #numeric
x <- c(T,F)                 #logic
x <- vector("numeric",length=10)  #also:complex,character
y <- c(1.7,"a")
y <- c(TRUE,"a")
z <- 0:8
class(z)
```

```
## [1] "integer"
```

```
as.numeric(z)
```

```
## [1] 0 1 2 3 4 5 6 7 8
```

```
as.character(z)
```

```
## [1] "0" "1" "2" "3" "4" "5" "6" "7" "8"
```

```
as.complex(z)
```

```
## [1] 0+0i 1+0i 2+0i 3+0i 4+0i 5+0i 6+0i 7+0i 8+0i
```

```
##list
x <- list(1,"a",1+8i,0.1,T)

##matrix
m <- matrix(nrow=3,ncol=4)
dim(m)
```

```
## [1] 3 4
```

```
attributes(m)
```

```
## $dim
## [1] 3 4
```

```
m <- matrix(1:12,nrow=3,ncol=4)
dim(m) <- c(2,6)
x <- 1:3
y <- 7:9
rbind(x,y)
```

```
##   [,1] [,2] [,3]
## x    1    2    3
## y    7    8    9
```

```
cbind(x,y)
```

```
##      x y
## [1,] 1 7
## [2,] 2 8
## [3,] 3 9
```

```
##factor
x <- factor(c("yes","yes","no","yes","no"))
table(x)
```

```
## x
##  no yes
##   2   3
```

```
unclass(x)
```

```
## [1] 2 2 1 2 1
## attr(,"levels")
## [1] "no"  "yes"
```

```
x <- factor(c("yes","yes","no","yes","no"),levels = c("yes","no"))
#yes is first level,no is second level

##Data Frame, read.table(), read.csv(), row.name, data.matrix()
x <- data.frame(foo = 1:4, bar = c(T, T, F, F))
nrow(x)
```

```
## [1] 4
```

```
ncol(x)
```

```
## [1] 2
```

```
##Names
x <- c(1, 2, 3, 4)
names(x)
```

```
## NULL
```

```
names(x) <- c('bar', 'jun', 'li', 'kk')
x
```

```
## bar jun  li  kk
##   1   2   3   4
```

```r
##Quiz1 dataset

##Detail of the dataset
data <- read.csv("hw1_data.csv")
str(data)
```

```
## 'data.frame':    153 obs. of  6 variables:
##  $ Ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...
##  $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
##  $ Wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
##  $ Temp   : int  67 72 74 62 56 66 65 59 61 69 ...
##  $ Month  : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ Day    : int  1 2 3 4 5 6 7 8 9 10 ...
```

```r
names(data)
```

```
## [1] "Ozone"   "Solar.R" "Wind"    "Temp"    "Month"   "Day"
```

```r
r2 <- file("hw1_data.csv", "r")
data2 <- readLines(r2,3)
close(r2)
nrow(data) #R can divide the first line (feature name)!!!
```

```
## [1] 153
```

```r
##14.Extract the last 2 rows of the data frame and print them to the console.
##What does the output look like?
data12 <- tail(data,2) #head,tail
print(data12)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 152    18     131  8.0   76     9  29
## 153    20     223 11.5   68     9  30
```

```r
##15.What is the value of Ozone in the 47th row?
print(data[47,1, drop = FALSE])
```

```
##    Ozone
## 47    21
```

```r
##16.How many missing values are in the Ozone column of this data frame?
sum(is.na(data$Ozone))
```

```
## [1] 37
```

```r
##17.What is the mean of the Ozone column in this dataset?
##Exclude missing values (coded as NA) from this calculation.
mean(data$Ozone[!is.na(data$Ozone)])
```

```
## [1] 42.12931
```

```
##18.Extract the subset of rows of the data frame where Ozone values are above 31 and Temp values are al
##What is the mean of Solar.R in this subset?

Solarset <- subset(data, Ozone >31 & Temp >90)
mean(Solarset$Solar.R[!is.na(Solarset$Solar.R)])
```

```
## [1] 212.8
```

```
##or another way
val <- complete.cases(data$Solar.R, data$Ozone, data$Temp)
mean(data$Solar.R[val & data$Ozone >31 & data$Temp >90])
```

```
## [1] 212.8
```

```
##19.What is the mean of "Temp" when "Month" is equal to 6?
mean(data$Temp[data$Month == 6])
```

```
## [1] 79.1
```

```
##20.What was the maximum ozone value in the month of May (i.e. Month is equal to 5)?
max(data$Ozone[data$Month == 5 & !is.na(data$Ozone)])
```

```
## [1] 115
```