Assignment 3

Pratishtha Deep

October 16, 2017

#QUE1: Import the data into R.  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

ForestFires = read.csv("C:/Users/pratishtha/Downloads/forestfires.csv", header = TRUE, sep = ',')  
head(ForestFires)

## X Y month day FFMC DMC DC ISI temp RH wind rain area  
## 1 7 5 mar fri 86.2 26.2 94.3 5.1 8.2 51 6.7 0.0 0  
## 2 7 4 oct tue 90.6 35.4 669.1 6.7 18.0 33 0.9 0.0 0  
## 3 7 4 oct sat 90.6 43.7 686.9 6.7 14.6 33 1.3 0.0 0  
## 4 8 6 mar fri 91.7 33.3 77.5 9.0 8.3 97 4.0 0.2 0  
## 5 8 6 mar sun 89.3 51.3 102.2 9.6 11.4 99 1.8 0.0 0  
## 6 8 6 aug sun 92.3 85.3 488.0 14.7 22.2 29 5.4 0.0 0

#####a. How many observations are there in the dataset?  
nrow(ForestFires)

## [1] 517

#####b. How many observations are there with a fire (i.e., area>0)  
nrow(subset(ForestFires, area>0))

## [1] 270

#####c. How many observations are there with a rain (i.e., rain>0)  
nrow(subset(ForestFires, rain > 0))

## [1] 8

#####d. How many observations are there with both a fire and a rain?  
nrow(ForestFires[ForestFires$area>0 & ForestFires$rain>0, ])

## [1] 2

#Que2: Show the columns month, day, area of the all the observations.  
head(ForestFires[, c('month', 'day', 'area')])

## month day area  
## 1 mar fri 0  
## 2 oct tue 0  
## 3 oct sat 0  
## 4 mar fri 0  
## 5 mar sun 0  
## 6 aug sun 0

#Que3: Show the columns month, day, area of the observations with a fire.  
head(ForestFires[ForestFires$area>0, c('month', 'day', 'area')])

## month day area  
## 139 jul tue 0.36  
## 140 sep tue 0.43  
## 141 sep mon 0.47  
## 142 aug wed 0.55  
## 143 aug fri 0.61  
## 144 jul sat 0.71

#Que4: How large are the five largest fires (i.e., having largest area)  
ForestFires\_df<-ForestFires[ForestFires$area>0, ]  
ForestFires\_arrange <- arrange(ForestFires\_df, desc(area))  
ForestFire\_largest<-(top\_n(ForestFires\_arrange, 5, area))  
ForestFire\_largest$area

## [1] 1090.84 746.28 278.53 212.88 200.94

####a. What are the corresponding month, temp, RH, wind, rain area?  
ForestFires\_df<-ForestFires[ForestFires$area>0, ]  
ForestFires\_arrange <- arrange(ForestFires\_df, desc(area))  
ForestFire\_largest<-(top\_n(ForestFires\_arrange, 5, area))  
head(ForestFire\_largest)

## X Y month day FFMC DMC DC ISI temp RH wind rain area  
## 1 6 5 sep sat 92.5 121.1 674.4 8.6 25.1 27 4.0 0 1090.84  
## 2 8 6 aug thu 94.8 222.4 698.6 13.9 27.5 27 4.9 0 746.28  
## 3 7 4 jul mon 89.2 103.9 431.6 6.4 22.6 57 4.9 0 278.53  
## 4 1 2 sep tue 91.0 129.5 692.6 7.0 18.8 40 2.2 0 212.88  
## 5 2 2 sep sat 92.5 121.1 674.4 8.6 18.2 46 1.8 0 200.94

####b. Reorder factor levels of month to be from Jan to Dec. Add one column to the data indicating whether a fire occurred for each observation (True for area >0 and False for area ==0)  
levels(ForestFires$month)

## [1] "apr" "aug" "dec" "feb" "jan" "jul" "jun" "mar" "may" "nov" "oct"  
## [12] "sep"

ForestFires$month = factor(ForestFires$month, levels = c('jan', 'feb', 'mar', 'apr', 'may', 'jun', 'jul', 'aug', 'sep', 'oct', 'nov', 'dec'))  
ForestFires$data<- ifelse(ForestFires$area>0, "TRUE", "FALSE")  
head(ForestFires)

## X Y month day FFMC DMC DC ISI temp RH wind rain area data  
## 1 7 5 mar fri 86.2 26.2 94.3 5.1 8.2 51 6.7 0.0 0 FALSE  
## 2 7 4 oct tue 90.6 35.4 669.1 6.7 18.0 33 0.9 0.0 0 FALSE  
## 3 7 4 oct sat 90.6 43.7 686.9 6.7 14.6 33 1.3 0.0 0 FALSE  
## 4 8 6 mar fri 91.7 33.3 77.5 9.0 8.3 97 4.0 0.2 0 FALSE  
## 5 8 6 mar sun 89.3 51.3 102.2 9.6 11.4 99 1.8 0.0 0 FALSE  
## 6 8 6 aug sun 92.3 85.3 488.0 14.7 22.2 29 5.4 0.0 0 FALSE

Or

Levels(df$month)

Nlevels

Df$month = factor(df$month, levels = month.abb)

Tolower(math.abb)