Assignment Basic Loading and Transformation R Markdown

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
# DATA 607 Week 1 Assignment - Loading Data into a Data Frame
# Student Name: Kalyan (Kalyanaraman Parthasarathy)

# Set working directory to D:\Data in which "agaricus-lepiota data.csv" is stored.
# The data set "expanded" is converted to a CSV file and loaded into the R program
setwd("D:/Data")

# Read telecomCalls.csv
dataset <- read.csv("agaricus-lepiota data.csv")

# Following are the general inspections of the data set
# Data Inspection
class(dataset)

# Number of columns in the data set
length(dataset)

# Number of rows in the data set
nrow(dataset)

dim(dataset)

# Structure
str(dataset)

# Naming the columns of the data frame - Used the file "agaricus-lepiota.names" to get
# the names for the columns except for the first column which I named it as "Edibleness"
colnames(dataset) <- c(
                    "Edibleness"
                    , "cap-shape"
                    , "cap-surface"
                    , "cap-color"
                    , "bruises?"
                    , "odor"
                    , "gill-attachment"
                    , "gill-spacing"
                    , "gill-size"
                    , "gill-color"
                    , "stalk-shape"
```

```r
                       , "stalk-root"
                       , "stalk-surface-above-ring"
                       , "stalk-surface-below-ring"
                       , "stalk-color-above-ring"
                       , "stalk-color-below-ring"
                       , "veil-type"
                       , "veil-color"
                       , "ring-number"
                       , "ring-type"
                       , "spore-print-color"
                       , "population"
                       , "habitat"
)


# Print the new column names
colnames(dataset)


# Create a new data frame with only first 5 columns
dataset_2 <- dataset[, c(1, 2, 3, 4, 5)]

# Data Inspection
class(dataset_2)

# Structure
str(dataset_2)

# Assign column names to the new data set
colnames(dataset_2) <- c(
                       "Edibleness"
                       , "shape"
                       , "surface"
                       , "color"
                       , "bruises"
                     )

# Print the column names
colnames(dataset_2)

# Change the data in "Edibleness" column according to the abbreviations below
# edible=e, non-edible=p
levels(dataset_2$Edibleness)[levels(dataset_2$Edibleness) == "e"] <- "edible"
levels(dataset_2$Edibleness)[levels(dataset_2$Edibleness) == "p"] <- "non-edible"

# Change the data in "shape" column according to the abbreviations below
# bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
levels(dataset_2$shape)[levels(dataset_2$shape) == "b"] <- "bell"
levels(dataset_2$shape)[levels(dataset_2$shape) == "c"] <- "conical"
levels(dataset_2$shape)[levels(dataset_2$shape) == "x"] <- "convex"
levels(dataset_2$shape)[levels(dataset_2$shape) == "f"] <- "flat"
levels(dataset_2$shape)[levels(dataset_2$shape) == "k"] <- "knobbed"
levels(dataset_2$shape)[levels(dataset_2$shape) == "s"] <- "sunken"

# Change the data in "surface" column according to the abbreviations below
```

```
# fibrous=f,grooves=g,scaly=y,smooth=s
levels(dataset_2$surface)[levels(dataset_2$surface) == "f"] <- "fibrous"
levels(dataset_2$surface)[levels(dataset_2$surface) == "g"] <- "grooves"
levels(dataset_2$surface)[levels(dataset_2$surface) == "y"] <- "scaly"
levels(dataset_2$surface)[levels(dataset_2$surface) == "s"] <- "smooth"

# Change the data in "color" column according to the abbreviations below
# brown=n,buff=b,cinnamon=c,gray=g,green=r,pink=p,purple=u,red=e,white=w,yellow=y
levels(dataset_2$color)[levels(dataset_2$color) == "n"] <- "brown"
levels(dataset_2$color)[levels(dataset_2$color) == "b"] <- "buff"
levels(dataset_2$color)[levels(dataset_2$color) == "c"] <- "cinnamon"
levels(dataset_2$color)[levels(dataset_2$color) == "g"] <- "gray"
levels(dataset_2$color)[levels(dataset_2$color) == "r"] <- "green"
levels(dataset_2$color)[levels(dataset_2$color) == "p"] <- "pink"
levels(dataset_2$color)[levels(dataset_2$color) == "u"] <- "purple"
levels(dataset_2$color)[levels(dataset_2$color) == "e"] <- "red"
levels(dataset_2$color)[levels(dataset_2$color) == "w"] <- "white"
levels(dataset_2$color)[levels(dataset_2$color) == "y"] <- "yellow"

# Change the data in "bruises" column according to the abbreviations below
# bruises=t,no=f
levels(dataset_2$bruises)[levels(dataset_2$bruises) == "t"] <- "bruises"
levels(dataset_2$bruises)[levels(dataset_2$bruises) == "f"] <- "no"

# Print data set 2 (5 columns data frame)
dataset_2
```