AsssignmentReport

Introduction

Since the significance of the knowledge about the edibility of mushroom, we believe it's necessary to extract meaningful infomation from the mushrooom dataset.This data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family.Each species is identified as edible or poisonous.This report mainly focus on the analysis of Mushroom dataset, including some static analysis in data exploration part, and association rules mining in data analysis part.

Data Exploration

Load the dataset at first.

```r
##load Mushroom dataset
url <- "./data/mushroom/agaricus-lepiota.data"
mushrooms <- read.csv(file = url, header = FALSE)
names(mushrooms) <- c("class", "cap-shape", "cap-surface",
                      "cap-color", "bruises", "odor", "gill-attachment", "gill-spacing",
                      "gill-size", "gill-color", "stalk-shape", "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")
```

Then check out the dimensionality of the dataset.

```r
dim(mushrooms)
```

```
## [1] 8124   23
```

And its column names.

```r
names(mushrooms)
```

```
##  [1] "class"                    "cap-shape"               
##  [3] "cap-surface"              "cap-color"               
##  [5] "bruises"                  "odor"                    
##  [7] "gill-attachment"          "gill-spacing"            
##  [9] "gill-size"                "gill-color"              
## [11] "stalk-shape"              "stalk-root"              
## [13] "stalk-surface-above-ring" "stalk-surface-below-ring"
## [15] "stalk-color-above-ring"   "stalk-color-below-ring"  
## [17] "veil-type"                "veil-color"              
## [19] "ring-number"              "ring-type"               
## [21] "spore-print-color"        "population"              
## [23] "habitat"
```

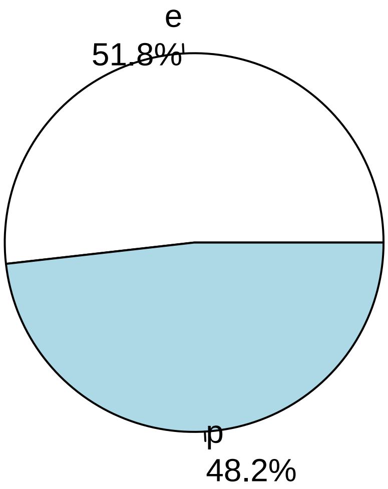
We can further explore these colunms.

```r
str(mushrooms)
```

```
## 'data.frame':    8124 obs. of  23 variables:
##  $ class                   : Factor w/ 2 levels "e","p": 2 1 1 2 1 1 1 1 2 1 ...
##  $ cap-shape               : Factor w/ 6 levels "b","c","f","k",..: 6 6 1 6 6 6 1 1 6 1 ...
##  $ cap-surface             : Factor w/ 4 levels "f","g","s","y": 3 3 3 4 3 4 3 4 4 3 ...
##  $ cap-color               : Factor w/ 10 levels "b","c","e","g",..: 5 10 9 9 4 10 9 9 9 10 ...
##  $ bruises                 : Factor w/ 2 levels "f","t": 2 2 2 2 1 2 2 2 2 2 ...
##  $ odor                    : Factor w/ 9 levels "a","c","f","l",..: 7 1 4 7 6 1 1 4 7 1 ...
##  $ gill-attachment         : Factor w/ 2 levels "a","f": 2 2 2 2 2 2 2 2 2 2 ...
##  $ gill-spacing            : Factor w/ 2 levels "c","w": 1 1 1 1 2 1 1 1 1 1 ...
##  $ gill-size               : Factor w/ 2 levels "b","n": 2 1 1 2 1 1 1 1 2 1 ...
##  $ gill-color              : Factor w/ 12 levels "b","e","g","h",..: 5 5 6 6 5 6 3 6 8 3 ...
##  $ stalk-shape             : Factor w/ 2 levels "e","t": 1 1 1 1 2 1 1 1 1 1 ...
##  $ stalk-root              : Factor w/ 5 levels "?","b","c","e",..: 4 3 3 4 4 3 3 3 4 3 ...
##  $ stalk-surface-above-ring: Factor w/ 4 levels "f","k","s","y": 3 3 3 3 3 3 3 3 3 3 ...
##  $ stalk-surface-below-ring: Factor w/ 4 levels "f","k","s","y": 3 3 3 3 3 3 3 3 3 3 ...
##  $ stalk-color-above-ring  : Factor w/ 9 levels "b","c","e","g",..: 8 8 8 8 8 8 8 8 8 8 ...
##  $ stalk-color-below-ring  : Factor w/ 9 levels "b","c","e","g",..: 8 8 8 8 8 8 8 8 8 8 ...
##  $ veil-type               : Factor w/ 1 level "p": 1 1 1 1 1 1 1 1 1 1 ...
##  $ veil-color              : Factor w/ 4 levels "n","o","w","y": 3 3 3 3 3 3 3 3 3 3 ...
##  $ ring-number             : Factor w/ 3 levels "n","o","t": 2 2 2 2 2 2 2 2 2 2 ...
##  $ ring-type               : Factor w/ 5 levels "e","f","l","n",..: 5 5 5 5 1 5 5 5 5 5 ...
##  $ spore-print-color       : Factor w/ 9 levels "b","h","k","n",..: 3 4 4 3 4 3 3 4 3 3 ...
##  $ population              : Factor w/ 6 levels "a","c","n","s",..: 4 3 3 4 1 3 3 4 5 4 ...
##  $ habitat                 : Factor w/ 7 levels "d","g","l","m",..: 6 2 4 6 2 2 4 4 2 4 ...
```

In order to know the split of between edible mushrooms and poisonous ones in the dataset, we can plot a pie chart.

```r
library(dplyr)
tab <- mushrooms$class %>% table()
precentages <- tab %>% prop.table() %>% round(3) * 100
txt <- paste0(names(tab), '\n', precentages, '%')
pie(tab, labels=txt)
```



Data Analysis:Association Rules Mining

Now we use Apriori to mine association rules from mushroom dataset and inspect the first 12 rules.

```
library(arules)
library(arulesViz)
rules <- apriori(mushrooms, control = list(verbose=F),
                 parameter = list(minlen=4, maxlen=5, confidence=1),
                 appearance = list(rhs=c("class=p", "class=e"),
                                   default="lhs"))
quality(rules) <- round(quality(rules), digits=3)
inspect(head(rules, 12))
```

```
##      lhs                             rhs       support confidence  lift count
## [1]  {gill-size=b,                                                          
##       gill-color=n,                                                         
##       ring-number=o}              => {class=e}   0.108          1 1.931   880
## [2]  {gill-attachment=f,                                                    
##       gill-size=b,                                                          
##       gill-color=n}               => {class=e}   0.100          1 1.931   816
## [3]  {gill-size=b,                                                          
##       gill-color=n,                                                         
##       veil-color=w}               => {class=e}   0.100          1 1.931   816
## [4]  {gill-size=b,                                                          
##       gill-color=n,                                                         
##       veil-type=p}                => {class=e}   0.108          1 1.931   880
## [5]  {odor=n,
##       stalk-root=e,
##       stalk-color-below-ring=w} => {class=e}   0.106          1 1.931   864
## [6]  {odor=n,
##       stalk-root=e,
##       stalk-color-above-ring=w} => {class=e}   0.106          1 1.931   864
## [7]  {bruises=f,
##       odor=n,
##       stalk-root=e}             => {class=e}   0.106          1 1.931   864
## [8]  {odor=n,
##       stalk-root=e,
##       ring-number=o}            => {class=e}   0.106          1 1.931   864
## [9]  {odor=n,
##       gill-attachment=f,
##       stalk-root=e}             => {class=e}   0.106          1 1.931   864
## [10] {odor=n,
##       stalk-root=e,
##       veil-color=w}             => {class=e}   0.106          1 1.931   864
## [11] {odor=n,
##       stalk-root=e,
##       veil-type=p}              => {class=e}   0.106          1 1.931   864
## [12] {bruises=f,
##       stalk-root=e,
##       stalk-color-below-ring=w} => {class=e}   0.106          1 1.931   864
```

Conclusion