omerozeren_HMW_1_Data607

```
library(ggplot2)
library(knitr)
download.file('https://archive.ics.uci.edu/ml/machine-learning-
databases/mushroom/agaricus-lepiota.data', 'agaricus-lepiota.data')
df <- read.table('agaricus-lepiota.data', sep=',', stringsAsFactors=FALSE)
```

Original Mushroom Data:

```
head(df)
```

```
##   V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20
## 1  p  x  s  n  t  p  f  c  n   k   e   e   s   s   w   w   p   w   o   p
## 2  e  x  s  y  t  a  f  c  b   k   e   c   s   s   w   w   p   w   o   p
## 3  e  b  s  w  t  l  f  c  b   n   e   c   s   s   w   w   p   w   o   p
## 4  p  x  y  w  t  p  f  c  n   n   e   e   s   s   w   w   p   w   o   p
## 5  e  x  s  g  f  n  f  w  b   k   t   e   s   s   w   w   p   w   o   e
## 6  e  x  y  y  t  a  f  c  b   n   e   c   s   s   w   w   p   w   o   p
##   V21 V22 V23
## 1   k   s   u
## 2   n   n   g
## 3   n   n   m
## 4   k   s   u
## 5   n   a   g
## 6   k   n   g
```

creating sunset of original data

```
df_subset <- subset(df,select=c(V1,V2,V3,V4,V23))
colnames(df_subset)<-
c("toxicity","cap_shape","cap_surface","cap_color","habitat")
head(df_subset)
```

```
##   toxicity cap_shape cap_surface cap_color habitat
## 1        p         x           s         n       u
## 2        e         x           s         y       g
## 3        e         b           s         w       m
## 4        p         x           y         w       u
## 5        e         x           s         g       g
## 6        e         x           y         y       g
```

```
df_subset[which(df_subset$toxicity=="e"),1] <- "edible"
df_subset[which(df_subset$toxicity=="p"),1] <- "poisonous"
df_subset[which(df_subset$cap_color=="n"),4] <- "brown"
df_subset[which(df_subset$cap_color=="g"),4] <- "gray"
df_subset[which(df_subset$cap_color=="e"),4] <- "red"
df_subset[which(df_subset$cap_color=="y"),4] <- "yellow"
df_subset[which(df_subset$cap_color=="w"),4] <- "white"
df_subset[which(df_subset$cap_color=="b"),4] <- "bluw"
```

```r
df_subset[which(df_subset$cap_color=="p"),4] <- "pink"
df_subset[which(df_subset$cap_shape=="b"),2] <- "bell"
df_subset[which(df_subset$cap_shape=="c"),2] <- "conical"
df_subset[which(df_subset$cap_shape=="f"),2] <- "flat"
df_subset[which(df_subset$cap_shape=="k"),2] <- "knobbed"
df_subset[which(df_subset$cap_shape=="s"),2] <- "sunken"
df_subset[which(df_subset$cap_shape=="x"),2] <- "convex"
df_subset[which(df_subset$cap_surface=="f"),3] <- "fibrous"
df_subset[which(df_subset$cap_surface=="g"),3] <- "grooves"
df_subset[which(df_subset$cap_surface=="s"),3] <- "scaly"
df_subset[which(df_subset$cap_surface=="y"),3] <- "smooth"
df_subset[which(df_subset$habitat=="g"),5] <- "grasses"
df_subset[which(df_subset$habitat=="l"),5] <- "leaves"
df_subset[which(df_subset$habitat=="m"),5] <- "meadows"
df_subset[which(df_subset$habitat=="p"),5] <- "paths"
df_subset[which(df_subset$habitat=="u"),5] <- "urban"
df_subset[which(df_subset$habitat=="w"),5] <- "waste"
df_subset[which(df_subset$habitat=="d"),5] <- "woods"
head(df_subset)
```

```
##    toxicity cap_shape cap_surface cap_color habitat
## 1 poisonous    convex       scaly     brown   urban
## 2    edible    convex       scaly    yellow grasses
## 3    edible      bell       scaly     white meadows
## 4 poisonous    convex      smooth     white   urban
## 5    edible    convex       scaly      gray grasses
## 6    edible    convex      smooth    yellow grasses
```

SUMMARY OF DATAFRAME

```r
summary(df_subset)
```

```
##    toxicity          cap_shape         cap_surface
##  Length:8124        Length:8124        Length:8124
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   cap_color           habitat
##  Length:8124        Length:8124
##  Class :character   Class :character
##  Mode  :character   Mode  :character
```