Load the mushroom data. From the Data Set Description, we should see 8,124 cases and 22 atributes in addition to the class (edible or poisonous).

```
mrdata <- read.csv(url("https://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/agaricus-lep
dim(mrdata)
```

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

Let's look at a couple records:

```
head(mrdata)
```

```
##   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
```

Create a data frame with a subset of the columns in the dataset.You should include the column that indicates edible or poisonous and three or four other columns.

```
subset <- mrdata[,c("V1","V2","V6","V11","V22","V23")]
head(subset)
```

```
##   V1 V2 V6 V11 V22 V23
## 1  p  x  p   e   s   u
## 2  e  x  a   e   n   g
## 3  e  b  l   e   n   m
## 4  p  x  p   e   s   u
## 5  e  x  n   t   a   g
## 6  e  x  a   e   n   g
```

```
dim(subset)
```

```
## [1] 8124    6
```

You should also add meaningful column names and replace the abbreviations used in the data-for example, in the appropriate column, "e" might become "edible."

```
names(subset) <- c("class","cap-shape","odor","stalk-shape","population","habitat")
names(subset)
```

```
## [1] "class"       "cap-shape"   "odor"        "stalk-shape" "population"
## [6] "habitat"
```

```
head(subset)
```

```
##   class cap-shape odor stalk-shape population habitat
## 1     p         x    p           e          s       u
## 2     e         x    a           e          n       g
## 3     e         b    l           e          n       m
## 4     p         x    p           e          s       u
## 5     e         x    n           t          a       g
## 6     e         x    a           e          n       g
```

Replace abbreviations:

```
subset$class <- ifelse(subset$class == 'e', 'edible','poisonous')
library(plyr)
subset$`cap-shape` <- revalue(subset$'cap-shape', c('b'='bell', 'c'='conical', 'x'='convex', 'f'='flat'
subset$odor <-revalue(subset$odor, c('a'='almond', 'l'='anise','c'='creosote', 'y'='fishy', 'f'='foul',
subset$`stalk-shape` <- revalue(subset$`stalk-shape`, c('e'='enlarging','t'='tapering'))
subset$population <- revalue(subset$population, c('a'='abundant','c'='clustered', 'n'='numerous', 's'='
subset$habitat <- revalue(subset$habitat, c('g'='grasses','l'='leaves','m'='meadows','p'='paths','u'='u
```

Look at this distrubition of levels to ensure revalued.

```
table(subset$`cap-shape`)
```

```
##
##    bell conical    flat knobbed  sunken   convex
##     452       4    3152     828      32     3656
```

```
table(subset$odor)
```

```
##
##   almond creosote     foul    anise    musty     none  pungent    spicy
##      400      192     2160      400       36     3528      256      576
##    fishy
##      576
```

```
table(subset$`stalk-shape`)
```

```
##
## enlarging  tapering
##      3516      4608
```

```
table(subset$population)
```

```
##
##  abundant clustered  numerous scattered   several  solitary
##       384       340       400      1248      4040      1712
```

```
table(subset$habitat)
```

```
##
##   woods grasses  leaves meadows   paths   urban   waste
##    3148    2148     832     292    1144     368     192
```

Save the original and subset datasets to Github

```
save(mrdata, file="mrdata.Rda")
save(subset, file="mrsubdata.Rda")
```