```
library(magrittr)
library(dplyr)
library(tibble)
library(purrr)
library(ggplot2)
```

The above code is used to use all the library that we need so we can perform EDA to our dataset more easier

```
mushrooms <- read.csv("C:/Users/Alyza/Downloads/mushrooms.csv")
```

We use the above code to read mushroom dataset and load it into R

```
dim(mushrooms)
```

We use the above code to show the number of row and column in mushroom dataset.

Output:

```
> dim(mushrooms)
[1] 8124   23
```

From the output, we know that the mushroom dataset has 23 columns and 8124 rows

```
colnames(mushrooms)
```

We use the above code to show the name of each column.

Output:

```
> colnames(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"
```

By reading throught the output, we know that the mushroom dataset is a dataset that has information about mushrooms characteristics, such as odor, cap.color, cap.shape, veil.type, bruises, and many more

```
head(mushrooms)
tail(mushrooms)
```

We use **head(mushrooms)** to show us the first six rows from mushroom dataset and we use **tail(mushrooms)** to show us the last six rows from our mushroom dataset.

Output:

```
> head(mushrooms)
  class cap.shape cap.surface cap.color bruises odor gill.attachment gill.spacing
1     p         x           s         n       t    p               f            c
2     e         x           s         y       t    a               f            c
3     e         b           s         w       t    l               f            c
4     p         x           y         w       t    p               f            c
5     e         x           s         g       f    n               f            w
6     e         x           y         y       t    a               f            c
  gill.size gill.color stalk.shape stalk.root stalk.surface.above.ring
1         n          k           e          e                        s
2         b          k           e          c                        s
3         b          n           e          c                        s
4         n          n           e          e                        s
5         b          k           t          e                        s
6         b          n           e          c                        s
  stalk.surface.below.ring stalk.color.above.ring stalk.color.below.ring veil.type
1                        s                      w                      w         p
2                        s                      w                      w         p
3                        s                      w                      w         p
4                        s                      w                      w         p
5                        s                      w                      w         p
6                        s                      w                      w         p
  veil.color ring.number ring.type spore.print.color population habitat
1          w           o         p                 k          s       u
2          w           o         p                 n          n       g
3          w           o         p                 n          n       m
4          w           o         p                 k          s       u
5          w           o         e                 n          a       g
6          w           o         p                 k          n       g

> tail(mushrooms)
     class cap.shape cap.surface cap.color bruises odor gill.attachment
8119     p         k           y         n       f    f               f
8120     e         k           s         n       f    n               a
8121     e         x           s         n       f    n               a
8122     e         f           s         n       f    n               a
8123     p         k           y         n       f    y               f
8124     e         x           s         n       f    n               a
     gill.spacing gill.size gill.color stalk.shape stalk.root
8119            c         n          b           t          ?
8120            c         b          y           e          ?
8121            c         b          y           e          ?
8122            c         b          n           e          ?
8123            c         n          b           t          ?
8124            c         b          y           e          ?
     stalk.surface.above.ring stalk.surface.below.ring stalk.color.above.ring
8119                        k                        s                      p
8120                        s                        s                      o
8121                        s                        s                      o
8122                        s                        s                      o
8123                        s                        k                      w
8124                        s                        s                      o
     stalk.color.below.ring veil.type veil.color ring.number ring.type
8119                      w         p          w           o         e
8120                      o         p          o           o         p
8121                      o         p          n           o         p
8122                      o         p          o           o         p
8123                      w         p          w           o         e
8124                      o         p          o           o         p
     spore.print.color population habitat
8119                 w          v       d
8120                 b          c       l
8121                 b          v       l
8122                 b          c       l
8123                 w          v       l
8124                 o          c       l
```

By reading through the output, we know that the output briefly informs us that our dataset only contains categorical data

`str(mushrooms)`

We use the above code to show us the structure of the mushroom dataset. It show us the number of row and column, the data types of each column, and also give us the example data for each column/variable.

Output:

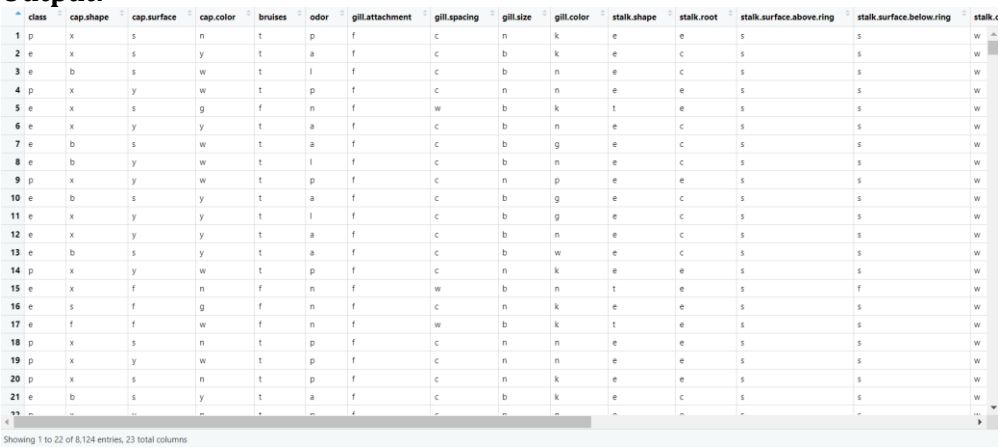
```
> str(mushrooms)
'data.frame':   8124 obs. of  23 variables:
 $ class                    : chr  "p" "e" "e" "p" ...
 $ cap.shape                : chr  "x" "x" "b" "x" ...
 $ cap.surface              : chr  "s" "s" "s" "y" ...
 $ cap.color                : chr  "n" "y" "w" "w" ...
 $ bruises                  : chr  "t" "t" "t" "t" ...
 $ odor                     : chr  "p" "a" "l" "p" ...
 $ gill.attachment          : chr  "f" "f" "f" "f" ...
 $ gill.spacing             : chr  "c" "c" "c" "c" ...
 $ gill.size                : chr  "n" "b" "b" "n" ...
 $ gill.color               : chr  "k" "k" "n" "n" ...
 $ stalk.shape              : chr  "e" "e" "e" "e" ...
 $ stalk.root               : chr  "e" "c" "c" "e" ...
 $ stalk.surface.above.ring: chr  "s" "s" "s" "s" ...
 $ stalk.surface.below.ring: chr  "s" "s" "s" "s" ...
 $ stalk.color.above.ring   : chr  "w" "w" "w" "w" ...
 $ stalk.color.below.ring   : chr  "w" "w" "w" "w" ...
 $ veil.type                : chr  "p" "p" "p" "p" ...
 $ veil.color               : chr  "w" "w" "w" "w" ...
 $ ring.number              : chr  "o" "o" "o" "o" ...
 $ ring.type                : chr  "p" "p" "p" "p" ...
 $ spore.print.color        : chr  "k" "n" "n" "k" ...
 $ population               : chr  "s" "n" "n" "s" ...
 $ habitat                  : chr  "u" "g" "m" "u" ...
```

By reading through it, we know that the variables that has “character” as it’s data type is all variables.

```
View(mushrooms)
```

We use the above code to open a new window that show us all the data from mushrooms dataset.

Output:



	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.c
1	p	x	s	n	t	p	f	c	n	k	e	e	s	s	w
2	e	x	s	y	t	a	f	c	b	k	e	c	s	s	w
3	e	b	s	w	t	l	f	c	b	n	e	c	s	s	w
4	p	x	y	w	t	p	f	c	n	n	e	e	s	s	w
5	e	x	s	g	f	n	f	w	b	k	t	e	s	s	w
6	e	x	y	y	t	a	f	c	b	n	e	c	s	s	w
7	e	b	s	w	t	a	f	c	b	g	e	c	s	s	w
8	e	b	y	w	t	l	f	c	b	n	e	c	s	s	w
9	p	x	y	w	t	p	f	c	n	p	e	e	s	s	w
10	e	b	s	y	t	a	f	c	b	g	e	c	s	s	w
11	e	x	y	y	t	l	f	c	b	g	e	c	s	s	w
12	e	x	y	y	t	a	f	c	b	n	e	c	s	s	w
13	e	b	s	y	t	a	f	c	b	w	e	c	s	s	w
14	p	x	y	w	t	p	f	c	n	k	e	e	s	s	w
15	e	x	f	n	f	n	f	w	b	n	t	e	s	f	w
16	e	s	f	g	f	n	f	c	n	k	e	e	s	s	w
17	e	f	f	w	f	n	f	w	b	k	t	e	s	s	w
18	p	x	s	n	t	p	f	c	n	n	e	e	s	s	w
19	p	x	y	w	t	p	f	c	n	n	e	e	s	s	w
20	p	x	s	n	t	p	f	c	n	k	e	e	s	s	w
21	e	b	s	y	t	a	f	c	b	k	e	c	s	s	w

Showing 1 to 22 of 8,124 entries, 23 total columns

By reading through the output, we know that the **mushrooms** dataset contains information about the characteristics of mushrooms, including

- class: edible=e, poisonous=p
- cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
- cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
- cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r,pink=p,purple=u,red=e,white=w,yellow=y
- bruises: bruises=t,no=f

- odor: almond=a,anise=l,creosote=c,fishy=y,foul=f,musty=m,none=n,pungent=p,spicy=s
- gill-attachment: attached=a,descending=d,free=f,notched=n
- gill-spacing: close=c,crowded=w,distant=d
- gill-size: broad=b,narrow=n
- gill-color: black=k,brown=n,buff=b,chocolate=h,gray=g, green=r,orange=o,pink=p,purple=u,red=e,white=w,yellow=y
- stalk-shape: enlarging=e,tapering=t
- stalk-root: bulbous=b,club=c,cup=u,equal=e,rhizomorphs=z,rooted=r,missing=?
- stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
- stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
- stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y
- stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y
- veil-type: partial=p,universal=u
- veil-color: brown=n,orange=o,white=w,yellow=y
- ring-number: none=n,one=o,two=t
- ring-type: cobwebby=c,evanescent=e,flaring=f,large=l,none=n,pendant=p,sheathing=s,zone=z
- spore-print-color: black=k,brown=n,buff=b,chocolate=h,green=r,orange=o,purple=u,white=w,yellow=y
- population: abundant=a,clustered=c,numerous=n,scattered=s,several=v,solitary=y
- habitat: grasses=g,leaves=l,meadows=m,paths=p,urban=u,waste=w,woods=d

By looking at the output we also know that null values in mushrooms dataset is represent with "?"

```
table(mushrooms$class)
```

We use the above code to counts the number of each unique value in class variable

Output:

```
> table(mushrooms$class)

   e    p
4208 3916
```

By reading through the output we know that
table(mushrooms$class) counts the number of instances of each unique value in variable/column "class", which represents whether the mushroom is edible or poisonous (edible=e, poisonous=p)

```
mushroom_table <- lapply(seq(from=2, to=ncol(mushrooms)),
                         function(x)     {table(mushrooms$class,
```

```
mushrooms[,x])})
names(mushroom_table) <- colnames(mushrooms)[2:ncol(mushrooms)]
for(i in 1:length(mushroom_table)) {
  print("====================================")
  print(names(mushroom_table)[i])
  print(mushroom_table[[i]])
}
```

we use the code above to create a table for each column/variable in our dataset (mushroom) except the first column (class). For each table, the rows will represent whether the column is edible or poisonous and the columns will represent all possible values in that variable. The resulting table will be stored in a list called mushroom table

Output:

```
[1] "===================================="
[1] "cap.shape"

      b    c    f    k    s    x
  e  404    0 1596  228   32 1948
  p   48    4 1556  600    0 1708
[1] "===================================="
[1] "cap.surface"

      f    g    s    y
  e 1560    0 1144 1504
  p  760    4 1412 1740
[1] "                                      "

[1] "===================================="
[1] "cap.color"

      b    c    e    g    n    p    r    u    w    y
  e   48   32  624 1032 1264   56   16   16  720  400
  p  120   12  876  808 1020   88    0    0  320  672
[1] "===================================="
[1] "bruises"

      f    t
  e 1456 2752
  p 3292  624
[1] "===================================="
[1] "odor"

      a    c    f    l    m    n    p    s    y
  e 400    0    0  400    0 3408    0    0    0
  p   0  192 2160    0   36  120  256  576  576
[1] "===================================="
[1] "gill.attachment"

      a    f
  e  192 4016
  p   18 3898
[1] "===================================="
[1] "gill.spacing"
```

```
      c    w
  e 3008 1200
  p 3804  112
[1] "======================================"
[1] "gill.size"

      b    n
  e 3920  288
  p 1692 2224
[1] "======================================"
[1] "gill.color"

      b    e    g    h    k    n    o    p    r    u    w    y
  e    0   96  248  204  344  936   64  852    0  444  956   64
  p 1728    0  504  528   64  112    0  640   24   48  246   22
[1] "======================================"
[1] "stalk.shape"

      e    t
  e 1616 2592
  p 1900 2016
[1] "======================================"
[1] "stalk.root"

      ?    b    c    e    r
  e  720 1920  512  864  192
  p 1760 1856   44  256    0
[1] "======================================"
[1] "stalk.surface.above.ring"

      f    k    s    y
  e  408  144 3640   16
  p  144 2228 1536    8
[1] "======================================"
[1] "stalk.surface.below.ring"

      f    k    s    y
  e  456  144 3400  208
  p  144 2160 1536   76
[1] "======================================"
[1] "stalk.color.above.ring"


      b    c    e    g    n    o    p    w    y
  e    0    0   96  576   16  192  576 2752    0
  p  432   36    0    0  432    0 1296 1712    8
[1] "======================================"
[1] "stalk.color.below.ring"

      b    c    e    g    n    o    p    w    y
  e    0    0   96  576   64  192  576 2704    0
  p  432   36    0    0  448    0 1296 1680   24
[1] "======================================"
[1] "veil.type"

      p
  e 4208
  p 3916
[1] "======================================"
[1] "veil.color"


      n    o    w    y
  e  96   96 4016    0
  p   0    0 3908    8
[1] "======================================"
[1] "ring.number"

      n    o    t
  e    0 3680  528
  p   36 3808   72
[1] "======================================"
[1] "ring.type"

      e    f    l    n    p
  e 1008   48    0    0 3152
  p 1768    0 1296   36  816
[1] "======================================"
[1] "spore.print.color"

      b    h    k    n    o    r    u    w    y
  e  48   48 1648 1744   48    0   48  576   48
  p   0 1584  224  224    0   72    0 1812    0
[1] "======================================"
[1] "population"
```

```
      a    c    n    s    v    y
  e 384  288  400  880 1192 1064
  p   0   52    0  368 2848  648
[1] "======================================"
[1] "habitat"

      d    g    l    m    p    u    w
  e 1880 1408  240  256  136   96  192
  p 1268  740  592   36 1008  272    0
```

as we look at the output above, we can compare the frequency of edible and poisonous mushrooms for each variable.

In the odor variable, we can see that each odor has a high probability of belonging to only one class. For example, if the odor is fishy, it is most likely poisonous (y in odor has e = 0 and p = 2160). The same pattern can be observed for other odors as well. Similarly, the stalk.surface.above.ring and stalk.surface.below.ring variables also show the same pattern. Therefore, it is likely that the odor, stalk.surface.above.ring, and stalk.surface.below.ring variables are highly correlated with the class variable.

Furthermore, from here we also know that the veil.type variable has the same value for all of its data, so we don't really need this data and we can simply drop this variable using the code below.

```
mushrooms <- mushrooms %>% select(- veil.type)
colnames(mushrooms)
```

we use the above code to select all columns from mushrooms dataset except veil.type and assigned back all the selecting columns into mushrooms using %>% operator, and then we check again all the columns in mushroom using colnames(mushrooms)

Output:

```
> mushrooms <- mushrooms %>% select(- veil.type)
> colnames(mushrooms)
 [1] "class"                    "cap.shape"                "cap.surface"              "cap.color"                "bruises"                  "odor"
 [7] "gill.attachment"          "gill.spacing"             "gill.size"                "gill.color"               "stalk.shape"              "stalk.root"
[13] "stalk.surface.above.ring" "stalk.surface.below.ring" "stalk.color.above.ring"  "stalk.color.below.ring"   "veil.color"               "ring.number"
[19] "ring.type"                "spore.print.color"        "population"               "habitat"
```

By looking through it we can see that now our dataset (mushrooms) doesn't has veil.type anymore.

```
mushrooms[mushrooms == "?"] <- NA

for(i in 1:ncol(mushrooms)){
  null_count <- sum(is.na(mushrooms[,i]))
  print(paste(colnames(mushrooms)[i],   "contains",   null_count,
"null data."))
}
```

We use the above code to change all the "?' elements in our dataset into NA, then we check and count all the missing values in every column with **`sum(is.na(mushrooms[,i]))`** and then we print the numbers of missing values in every columns.

Output:

```
> mushrooms[mushrooms == "?"] <- NA
> for(i in 1:ncol(mushrooms)){
+   null_count <- sum(is.na(mushrooms[,i]))
+   print(paste(colnames(mushrooms)[i], "contains", null_count, "null data."))
+ }
[1] "class contains 0 null data."
[1] "cap.shape contains 0 null data."
[1] "cap.surface contains 0 null data."
[1] "cap.color contains 0 null data."
[1] "bruises contains 0 null data."
[1] "odor contains 0 null data."
[1] "gill.attachment contains 0 null data."
[1] "gill.spacing contains 0 null data."
[1] "gill.size contains 0 null data."
[1] "gill.color contains 0 null data."
[1] "stalk.shape contains 0 null data."
[1] "stalk.root contains 2480 null data."
[1] "stalk.surface.above.ring contains 0 null data."
[1] "stalk.surface.below.ring contains 0 null data."
[1] "stalk.color.above.ring contains 0 null data."
[1] "stalk.color.below.ring contains 0 null data."
[1] "veil.color contains 0 null data."
[1] "ring.number contains 0 null data."
[1] "ring.type contains 0 null data."
[1] "spore.print.color contains 0 null data."
[1] "population contains 0 null data."
[1] "habitat contains 0 null data."
```

As we see the output above, we can see that our dataset contains 2480 null data in column stalk.root

```
mushrooms <- mushrooms[, colSums(is.na(mushrooms)) == 0]

for(i in 1:ncol(mushrooms)){
  null_count <- sum(is.na(mushrooms[,i]))
   print(paste(colnames(mushrooms)[i],  "contains",  null_count,
"null data."))
}
```

We use the above code removes any columns in the **mushrooms** data frame that contain missing values, then we check and count all the missing values in every column with **`sum(is.na(mushrooms[,i]))`** and then we print the numbers of missing values in every columns.

Output:

```
> mushrooms <- mushrooms[, colSums(is.na(mushrooms)) == 0]
>
> for(i in 1:ncol(mushrooms)){
+   null_count <- sum(is.na(mushrooms[,i]))
+   print(paste(colnames(mushrooms)[i], "contains", null_count, "null data."))
+ }
[1] "class contains 0 null data."
[1] "cap.shape contains 0 null data."
[1] "cap.surface contains 0 null data."
[1] "cap.color contains 0 null data."
[1] "bruises contains 0 null data."
[1] "odor contains 0 null data."
[1] "gill.attachment contains 0 null data."
[1] "gill.spacing contains 0 null data."
[1] "gill.size contains 0 null data."
[1] "gill.color contains 0 null data."
[1] "stalk.shape contains 0 null data."
[1] "stalk.surface.above.ring contains 0 null data."
[1] "stalk.surface.below.ring contains 0 null data."
[1] "stalk.color.above.ring contains 0 null data."
[1] "stalk.color.below.ring contains 0 null data."
[1] "veil.color contains 0 null data."
[1] "ring.number contains 0 null data."
[1] "ring.type contains 0 null data."
[1] "spore.print.color contains 0 null data."
[1] "population contains 0 null data."
[1] "habitat contains 0 null data."
```

As we see the output above, we can see that our dataset doesn't contains null data anymore.

```
duplicated_rows <- duplicated(mushrooms)
sum(duplicated_rows)
```

We use the above code to count all the duplicated rows in mushroom dataset. **Duplicated(mushrooms)** will return logical vector whether the row is duplicated or not then we assigned the vector into **duplicated_rows.** With **sum()** it will give us the number of rows that is duplicated.

Output:

```
> duplicated_rows <- duplicated(mushrooms)
> sum(duplicated_rows)
[1] 0
```

By reading through the output, we know that the mushroom dataset doesn't has duplicated rows.

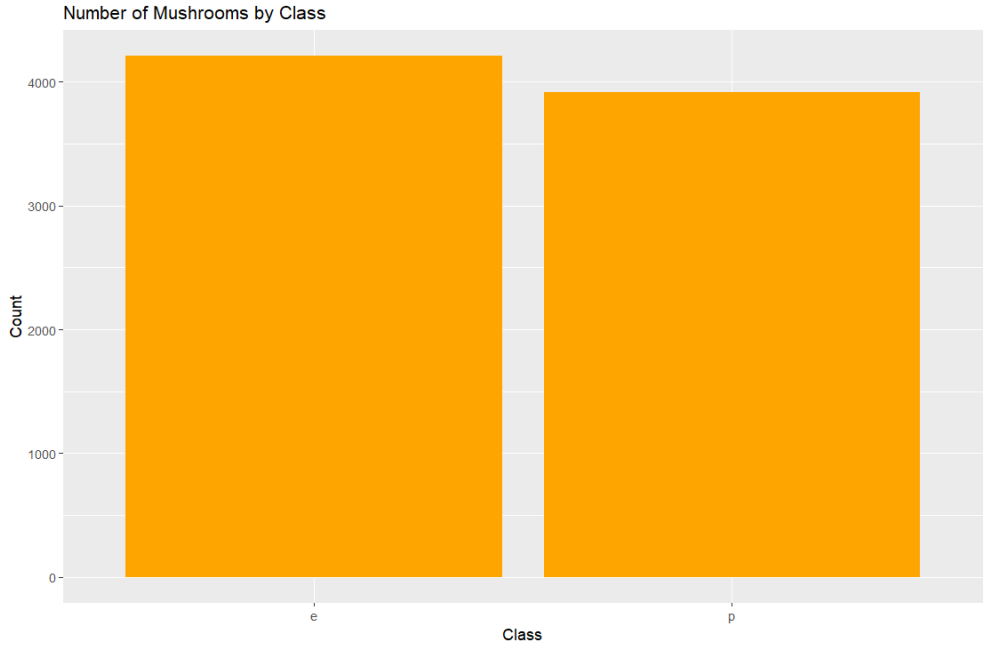
```
counts <- table(mushrooms$class)

ggplot(data = data.frame(Class = names(counts), Count = as.vector(counts)), aes(x = Class, y = Count)) +
  geom_bar(stat = "identity", fill = "orange") +
  labs(x = "Class", y = "Count", title = "Number of Mushrooms by Class")
```

we use this code to create a bar plot using ggplot() to show us the comparison of mushroom that is poisonous and mushroom that is edible.

counts <- table(mushrooms$class) used to count and assign the number of edible and poisonous mushroom into a table.

Output:



By looking at the output we can see that the number of edible mushrooms are more

bigger than the number of poisonous mushroom.

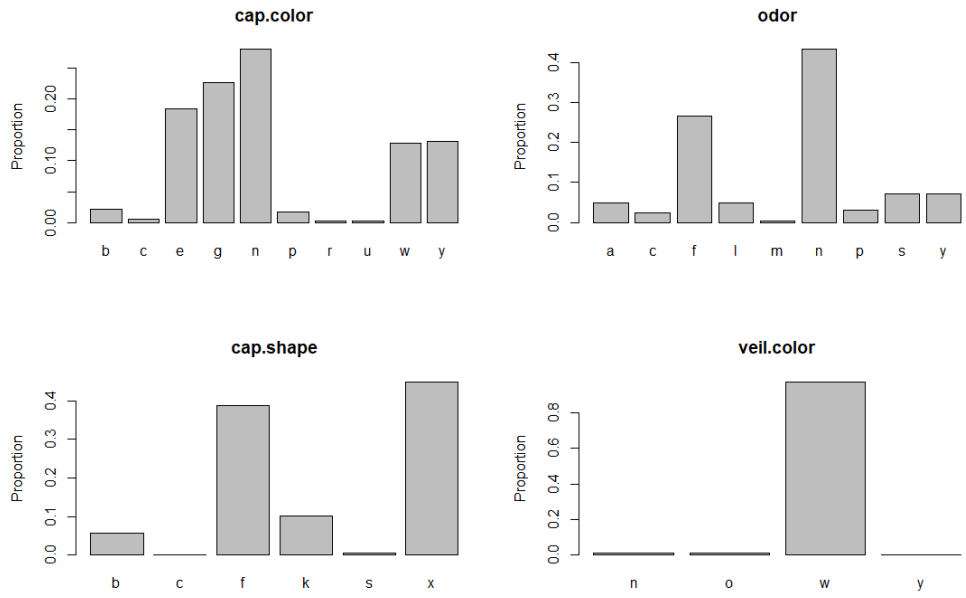
```
columns <- c('cap.color', 'odor', 'cap.shape', 'veil.color')
par(mfrow = c(2,2))
for(i in columns){
  freq_table <- prop.table(table(mushrooms[,i]))
  barplot(freq_table, main=i, ylab="Proportion")
}
```

we use this code to see the distribution of values in column cap.color, odor, cap.shape, and veil.color. This code will plot the distribution in the form of barplot.

freq_table <- prop.table(table(mushrooms[,i])) used to create a frequency table of each value in column cap.color, odor, cap.shape, and veil.color.

barplot(freq_table, main=i, ylab="Proportion") used to create the barplot using the data in freq_table

Output:



As we observe the plots above, we can see the proportion of each value in the variables cap.color, odor, cap.shape, and veil.color.

For cap.color, the top three colors in our dataset are:

- N (brown)
- G (gray)
- E (red)

For odor, the top two odors that dominate our dataset are:

- N (none)
- F (foul)

For cap.shape, the top three shapes that dominate our dataset are:

- X (convex)
- F (flat)
- K (knobbed)

For veil.color, white (w) dominates all other colors.

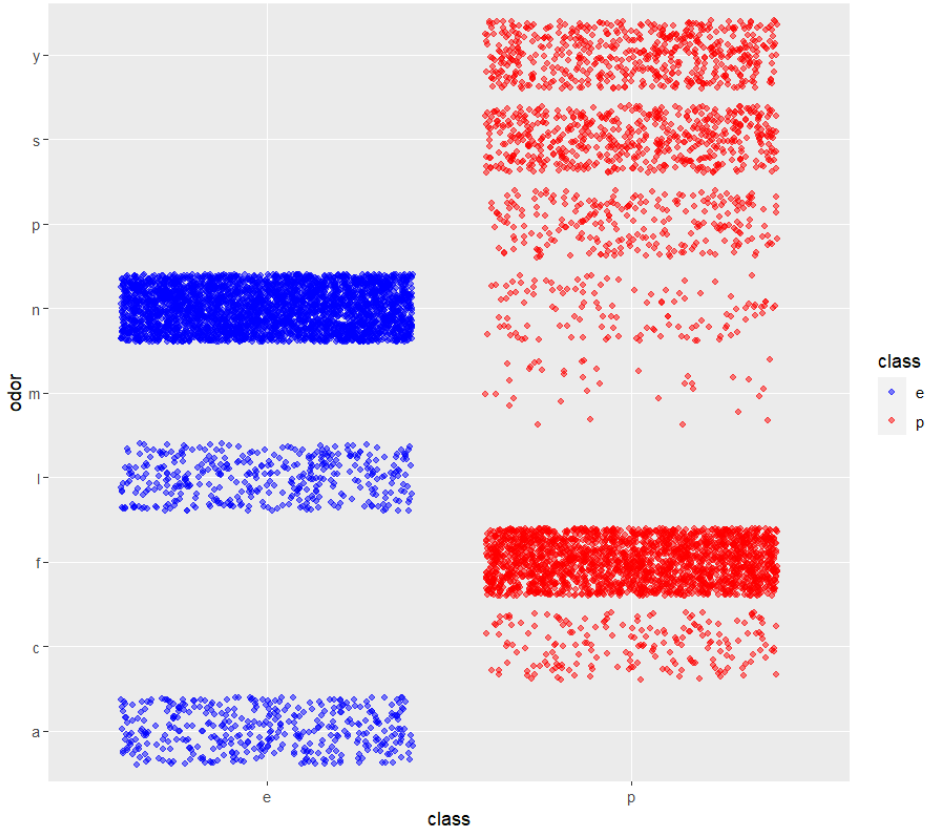
```
dataVis <- function(data, x, y, col) {
  x <- rlang::sym(x)
  y <- rlang::sym(y)
  col <- rlang::sym(col)

  ggplot(data = data, aes(x = !!x , y = !!y , col = !!col)) +
    geom_jitter(alpha = 0.5) +
    scale_color_manual(values = c("blue", "red"))
}

set.seed(1)
dataVis(data = mushrooms, x = 'class', y = 'odor', col = 'class')
```

We used the code above to create a scatter plot using ggplot2, which shows us the relationship between the variables "class" and "odor" in mushrooms.

Output:



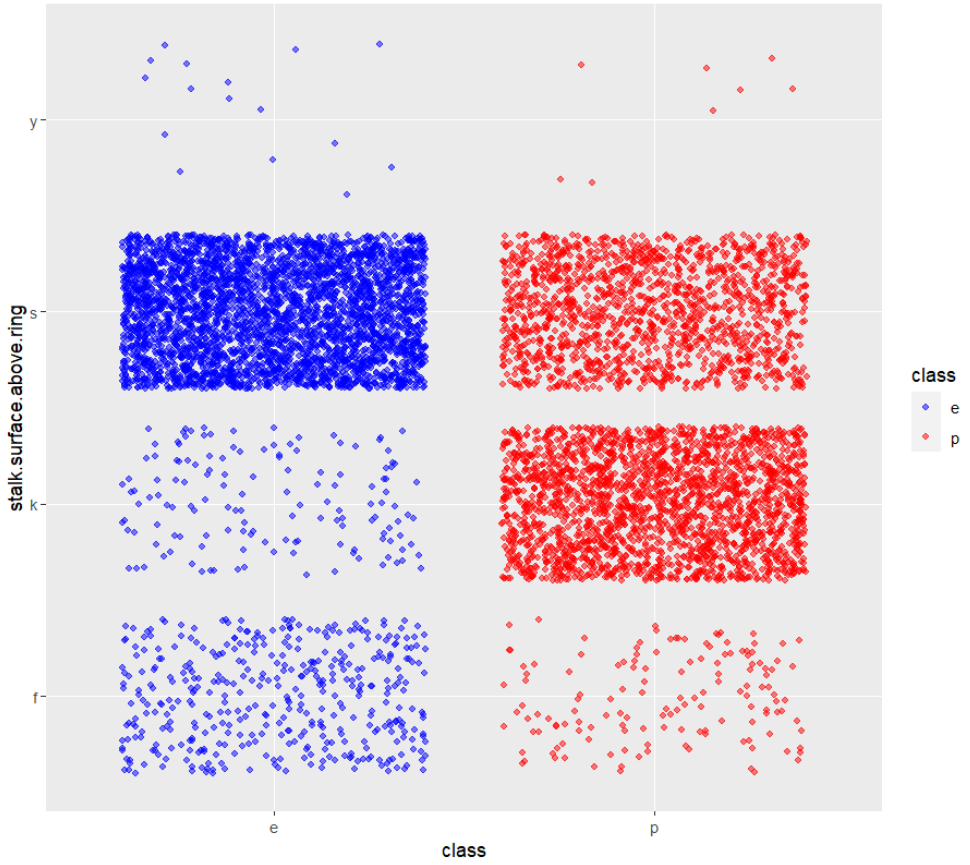
As we observe the scatter plot, we can see that if the mushroom odor is fishy, spicy, pungent, foul, or like creosote, it is more likely to be poisonous. On the other hand, if the mushroom odor is none, almond, or anise, it is more likely to be edible. Therefore, we can conclude that odor is a variable that is highly correlated with the

class (edible or poisonous) of mushrooms.

```
dataVis(data = mushrooms, x = 'class', y = 'stalk.surface.above.ring', col = 'class')
```

We used the code above to create a scatter plot using ggplot2, which shows us the relationship between the variables "class" and "stalk.surface.above.ring" in mushrooms.

Output:

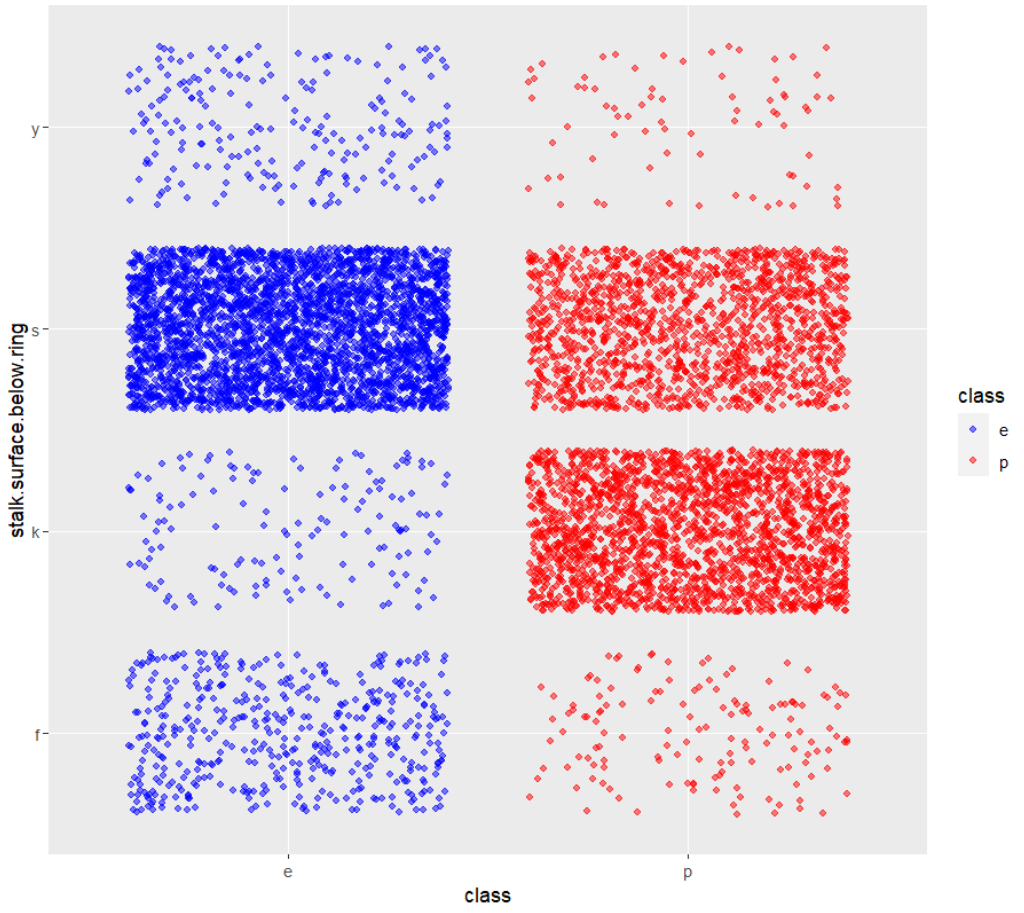


As we observe the scatter plot, we can see that if the stalk surface above ring is scaly or fibrous, it is more likely to be edible. On the other hand, if the stalk surface above ring is silky, it is more likely to be poisonous. However, for the stalk surface above ring that is smooth, it is somewhat difficult to see its distribution as the difference between the proportion of edible and poisonous mushrooms is not significant. Therefore, we can conclude that stalk surface above ring is a variable that is fairly correlated with the class (edible or poisonous) of mushrooms.

```
dataVis(data = mushrooms, x = 'class', y = 'stalk.surface.below.ring', col = 'class')
```

We used the code above to create a scatter plot using ggplot2, which shows us the relationship between the variables "class" and "stalk.surface.below.ring" in mushrooms.

Output:



As we observe the scatter plot, we can see that if the stalk surface below ring is scaly or fibrous, it is more likely to be edible. On the other hand, if the stalk surface below ring is silky, it is more likely to be poisonous. However, for the stalk surface below ring that is smooth, it is somewhat difficult to see its distribution as the difference between the proportion of edible and poisonous mushrooms is not significant. Therefore, we can conclude that stalk surface below ring is a variable that is fairly correlated with the class (edible or poisonous) of mushrooms.

SUMMARY

- **mushrooms** dataset is a dataset that displays characteristics about mushrooms such as **cap.color, odor, stalk.surface.above.ring,** and others**. This dataset is used to classify mushrooms into 2 classes, namely edible** or **poisonous** based on the characteristics given in the dataset
- Our dataset consists of 8124 rows and 23 columns
- All variables in our dataset are categorical data (data type = char)
- Missing value in our dataset represent by "?"
- Our dataset has 2480 missing value in stalk.root column and doesn't has any duplicated rows
- In our dataset, there are more edible mushrooms than poisonous ones
- The variable that is highly correlated with class is odor, while the stalk surface above and below the ring is fairly correlated with class