The Mushroom Study – Descriptive Analysis
Group 3: Chathu, Curtis, Tammy

The descriptive analysis starts with data cleaning and exploration. The analysis follows by dummy variable transformation and dimension reduction such as removing low correlation and high multicollinearity. Main descriptive analysis including Ridge regression, Linear Modelling, and ANOVA. Many forms of visualizations are included such as bar charts, correlation corrplot, boxplot, histograms, and etc.

DATA CLEANING AND EXPLORATION

The data contains 8,124 observations of individual mushroom data with 23 variables (Figure 1, Figure 2, Figure 3). Ring_number and poisonous are integers and the rest 21 variables are all categorical. The data is quite clean as there are no missing values (Figure 4). Since the poisonous is the target, the ring_number is the only numeric variable with data ranges in 0, 1 and 2 so there is no obvious outlier in the dataset (Figure 6).

```
head(df)
```

Description: df [6 × 23]

	cap_shape <chr>	cap_surface <chr>	cap_color <chr>	bruises <chr>	odor <chr>	gill_attachment <chr>	gill_spacing <chr>	gill_size <chr>	
1	undefined	smooth	brown	True	pungent	free	close	narrow	▸
2	undefined	smooth	yellow	True	almond	free	close	broad	
3	bell	smooth	white	True	anise	free	close	broad	
4	undefined	scaly	white	True	pungent	free	close	narrow	
5	undefined	smooth	grey	False	none	free	crowded	broad	
6	undefined	scaly	yellow	True	almond	free	close	broad	

6 rows | 1–9 of 23 columns

Figure 1 Data Preview

```
str(df) # Classification dataset of 8124 records and 23 variables, 21 chr and 2 int
```

```
'data.frame':    8124 obs. of  23 variables:
 $ cap_shape               : chr  "undefined" "undefined" "bell" "undefined" ...
 $ cap_surface             : chr  "smooth" "smooth" "smooth" "scaly" ...
 $ cap_color               : chr  "brown" "yellow" "white" "white" ...
 $ bruises                 : chr  "True" "True" "True" "True" ...
 $ odor                    : chr  "pungent" "almond" "anise" "pungent" ...
 $ gill_attachment         : chr  "free" "free" "free" "free" ...
 $ gill_spacing            : chr  "close" "close" "close" "close" ...
 $ gill_size               : chr  "narrow" "broad" "broad" "narrow" ...
 $ gill_color              : chr  "black" "black" "brown" "brown" ...
 $ stalk_shape             : chr  "enlarging" "enlarging" "enlarging" "enlarging" ...
 $ stalk_root              : chr  "equal" "club" "club" "equal" ...
 $ stalk_surface_above_ring: chr  "smooth" "smooth" "smooth" "smooth" ...
 $ stalk_surface_below_ring: chr  "smooth" "smooth" "smooth" "smooth" ...
 $ stalk_color_above_ring  : chr  "white" "white" "white" "white" ...
 $ stalk_color_below_ring  : chr  "white" "white" "white" "white" ...
 $ veil_type               : chr  "partial" "partial" "partial" "partial" ...
 $ veil_color              : chr  "white" "white" "white" "white" ...
 $ ring_number             : int  1 1 1 1 1 1 1 1 1 1 ...
 $ ring_type               : chr  "pendant" "pendant" "pendant" "pendant" ...
 $ spore_point_color       : chr  "black" "brown" "brown" "black" ...
 $ population              : chr  "scattered" "numerous" "numerous" "scattered" ...
 $ habitat                 : chr  "urban" "grass" "meadow" "urban" ...
 $ poisonous               : int  1 0 0 1 0 0 0 0 1 0 ...
```

Figure 2 Data Structure

```
summary(df)
```

2

```
   cap_shape          cap_surface         cap_color           bruises             odor
 Length:8124        Length:8124         Length:8124         Length:8124         Length:8124
 Class :character   Class :character    Class :character    Class :character    Class :character
 Mode  :character   Mode  :character    Mode  :character    Mode  :character    Mode  :character


 gill_attachment    gill_spacing        gill_size           gill_color          stalk_shape
 Length:8124        Length:8124         Length:8124         Length:8124         Length:8124
 Class :character   Class :character    Class :character    Class :character    Class :character
 Mode  :character   Mode  :character    Mode  :character    Mode  :character    Mode  :character


  stalk_root        stalk_surface_above_ring stalk_surface_below_ring stalk_color_above_ring
 Length:8124        Length:8124              Length:8124              Length:8124
 Class :character   Class :character         Class :character         Class :character
 Mode  :character   Mode  :character         Mode  :character         Mode  :character


 stalk_color_below_ring  veil_type          veil_color          ring_number       ring_type
 Length:8124            Length:8124         Length:8124        Min.   :0.000     Length:8124
 Class :character       Class :character    Class :character   1st Qu.:1.000     Class :character
 Mode  :character       Mode  :character    Mode  :character   Median :1.000     Mode  :character
                                                               Mean   :1.069
                                                               3rd Qu.:1.000
                                                               Max.   :2.000
 spore_point_color   population          habitat            poisonous
 Length:8124        Length:8124         Length:8124        Min.   :0.000
 Class :character   Class :character    Class :character   1st Qu.:0.000
 Mode  :character   Mode  :character    Mode  :character   Median :0.000
                                                           Mean   :0.482
                                                           3rd Qu.:1.000
                                                           Max.   :1.000
```

Figure 3 Data Summary

```
sum(is.na(df)) # Check for N/A

[1] 0
```

Figure 4 Missing Values

```
boxplot(df$ring_number) # Check for outliers: NONE Ring_number is the only numeric
```
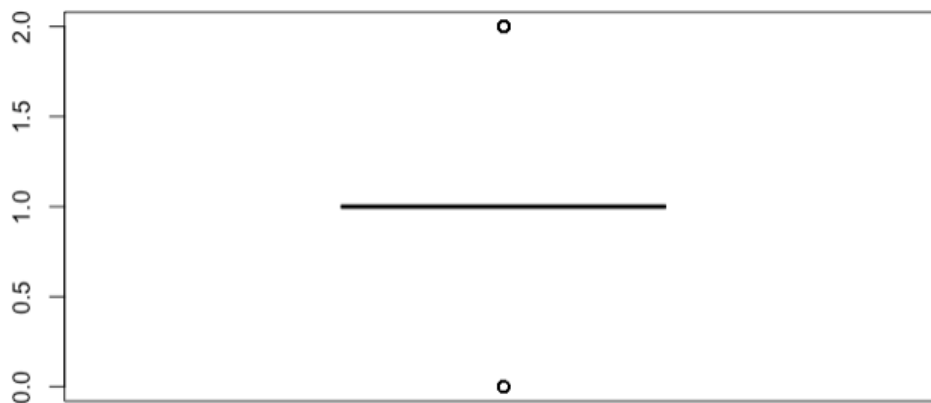

Figure 5 Outlier Check for Numeric Variable

DESCRIPTIVE ANALYSIS

The data contains 3,916 poisonous mushrooms and 4,208 non-poisonous mushrooms out of 8124 records which is a 48% poisonous rate (Figure 6).

```
count_class <- table(df$poisonous)
count_class
prop.table(count_class)*100

   0    1
4208 3916

        0        1
51.79714 48.20286
```

Figure 6 Dependent Variable Distribution

There are various attributes attached to the mushroom cap, bruises, odor, gill, stalk, veil, ring, spore, population and habitat and there are factors in each attribute (Figure 7).

```
[1] "cap_shape : 6"
[1] "cap_surface : 4"
[1] "cap_color : 10"
[1] "bruises : 2"
[1] "odor : 9"
[1] "gill_attachment : 2"
[1] "gill_spacing : 2"
[1] "gill_size : 2"
[1] "gill_color : 12"
[1] "stalk_shape : 2"
[1] "stalk_root : 5"
[1] "stalk_surface_above_ring : 4"
[1] "stalk_surface_below_ring : 4"
[1] "stalk_color_above_ring : 9"
[1] "stalk_color_below_ring : 9"
[1] "veil_type : 1"
[1] "veil_color : 4"
[1] "ring_number : 3"
[1] "ring_type : 5"
[1] "spore_point_color : 9"
[1] "population : 6"
[1] "habitat : 7"
```

Figure 7 List of Categorical Factors and Attributes in the Mushroom Dataset

We also visually explored correlation relationships between each variable and the target variable poisonous using ggplots. Early hypothesis on relationships can be drawn on variables such as **cap_surface, Bruises, Odor, gill_spacing, gill_size, gill_color, stalk_root, stalk_surface, stalk_color, ring_type, population and waste** and the relationships are examined further in the next section. Multiple visual representations are displayed such as bar charts (Figure 8), scatterplots (Figure 9) and bar plots (Figure 10).

Bar charts:

Figure 8 Relationships between independent variables and Poisonous in bar charts
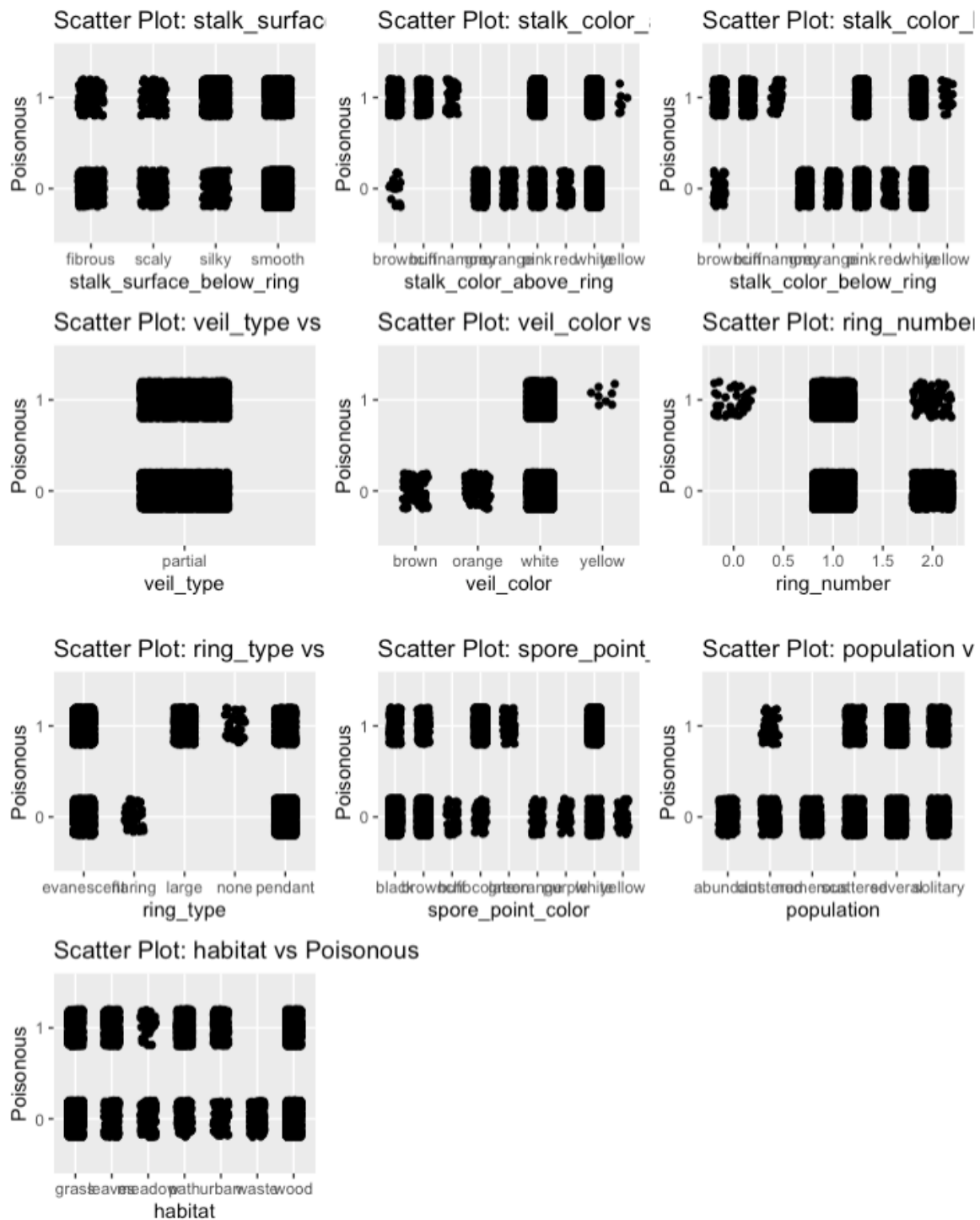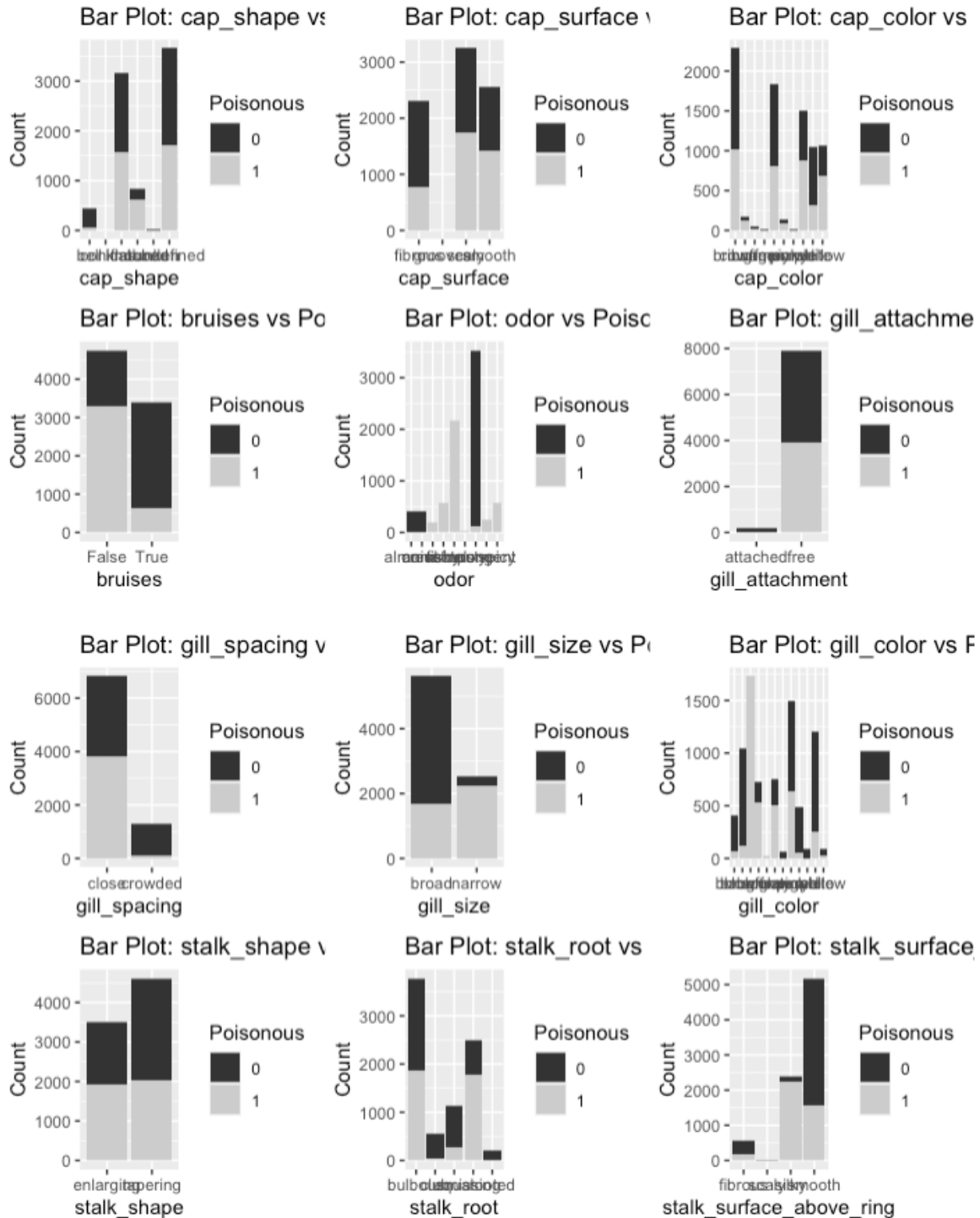
Scatterplots:
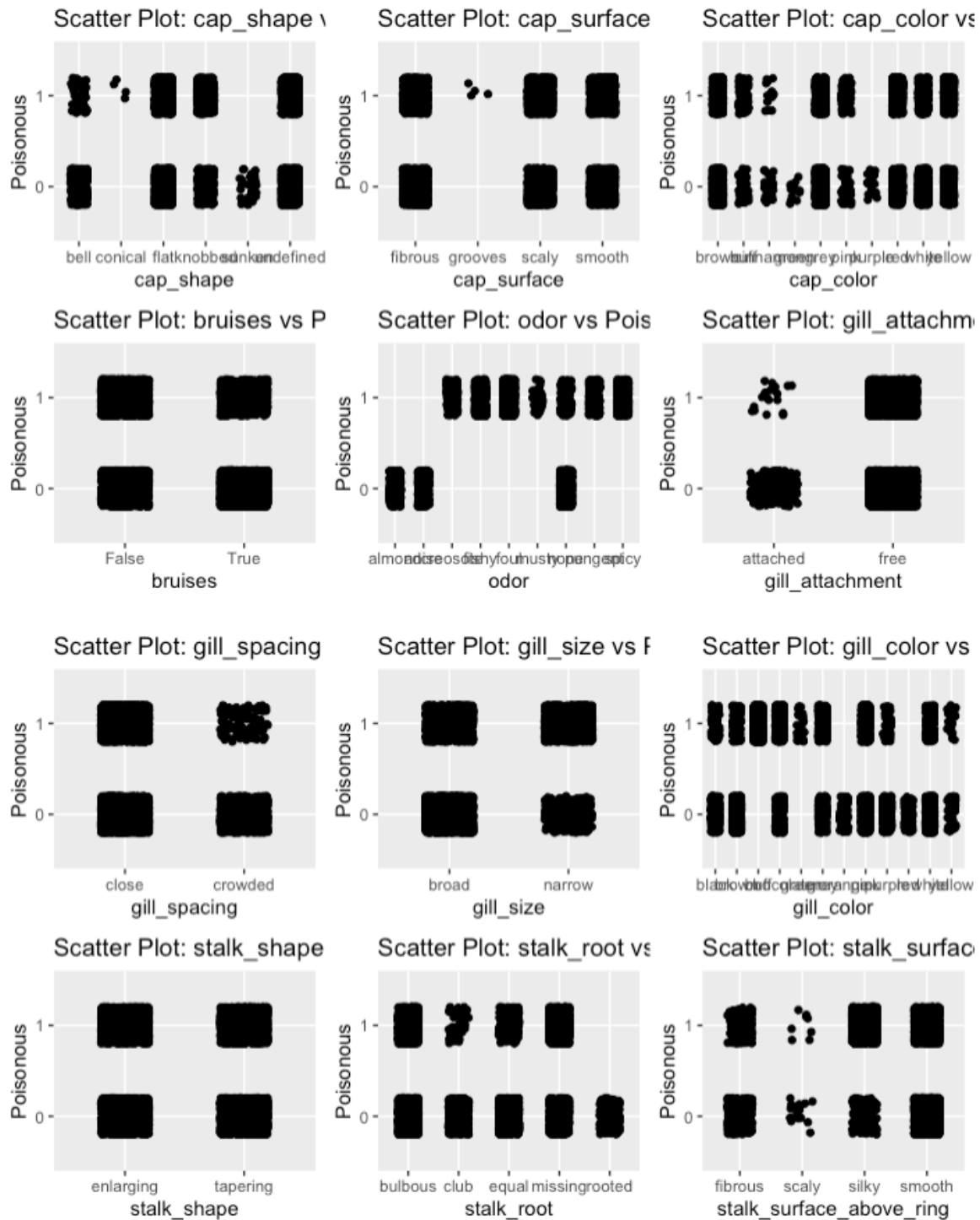
Figure 9 Relationships between independent variables and Poisonous in scatterplots
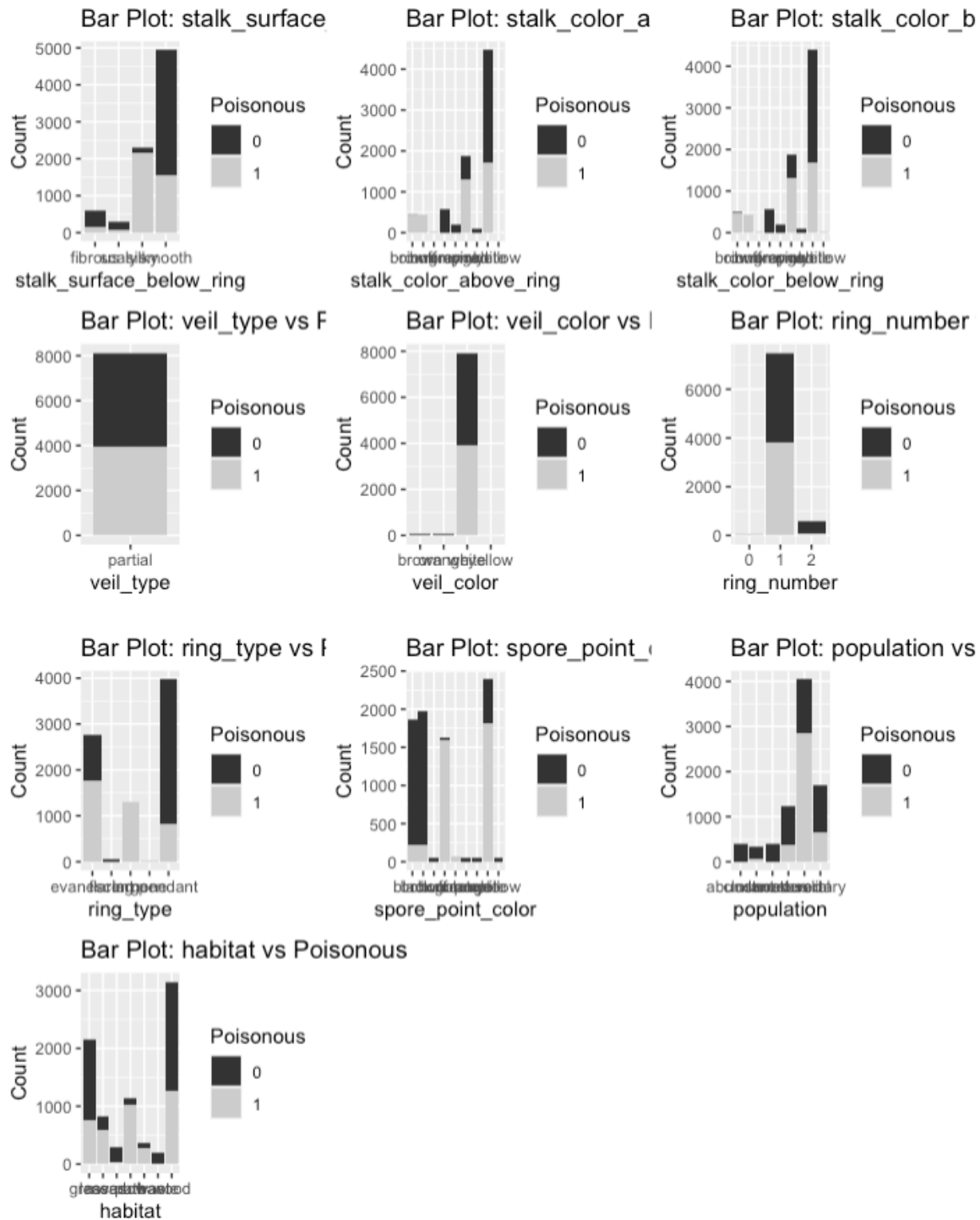
Bar plots:

Figure 10 Relationships between independent variables and Poisonous in bar plots

DUMMY VARIABLE TRANSFORMATION

To perform analysis on the mushroom dataset where all independent variables can be treated as categorical data, it is important to perform dummy variable transformation. Therefore, 22 independent variables are transformed into 117 independent variables. For example, bruises variable in Figure 10 can be transformed into two variables after the transformation as bruises_yes and bruises_no.

DIMENSION REDUCTION

- 80 out of 117 independent variables remained after removing variables have low correlations (<0.1) with target variable (Figure 11)
- Remove multicollinearity, independent variables that have high correlation with each other (Figure 12)

```
# Low correlation with target variable
dv_factor <- as.factor(df$poisonous) # dv in factor
dv_int <- df$poisonous # dv in num
cor_index <- as.vector(cor(df[, -2], dv_int))
plot(density(cor_index)) # Shows correlation density between iv and dv

threshold <- 0.1 # Set correlation threshold and remove variables with lower correlation
selected_indices <- which(abs(cor_index) > threshold)
selected_variables <- names(df)[selected_indices]
df1 <- cbind(df[, c(selected_variables)], dv_factor)
```

density.default(x = cor_index)

N = 116   Bandwidth = 0.0761
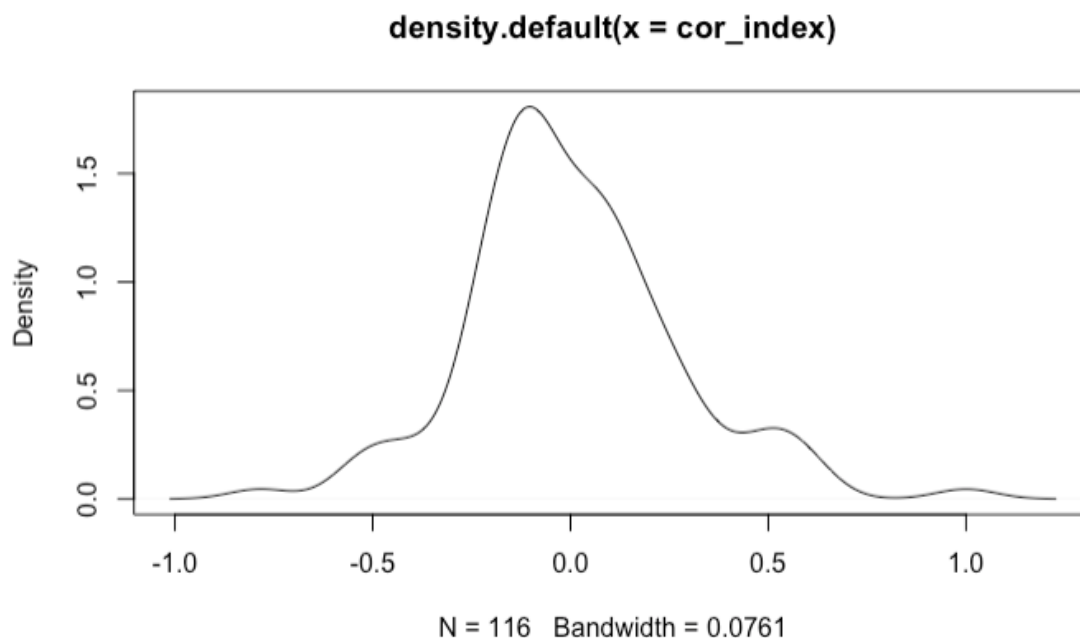
```
'data.frame':   8124 obs. of  80 variables:
```

Figure 11 IVs of Low Correlation with Target Variable

```
# Reorder columns so that they are ranked by number of unique values from highest to lowest.
rm(num_of_unique)
num_of_unique <- c()

for (x in 1:(length(df1-1))) {
  num_of_unique = append(num_of_unique,length(unique(df1[, x])))
}
rkk <- rank(-num_of_unique, ties.method= "first")
rkk<- append(rkk, length(df1))

rm(reorder_index)
reorder_index <- c()
for (x in 1:(length(df1))) {
  reorder_index = append(reorder_index,which(rkk == x))
}
df1_1 <- df1[, reorder_index]


# In each highly correlated pair, remove the second element
cor_matrix <- cor(df1_1[, -ncol(df1_1)]) # IV correlation matrix
cor_upper <- cor_matrix * upper.tri(cor_matrix, diag = FALSE) # Matrix of 0 and correlation

index <- apply(cor_upper, 1, function(x) paste(colnames(cor_upper)[which(abs(x) > 0.9)], collapse = ", "))
# Store IV names if correlation absolute value is larger than certain threshold
elements <- unique(unlist(strsplit(index, split = ", ")))
# Split names in one string to chr vector and remove duplicated IV names
#elements

cols_to_keep <- setdiff(names(df1_1), elements)
df2 <- df1_1[, cols_to_keep] # IVs remained at 0.9 threshold
```

Figure 12 IVs of High Correlation Between Each Other

FEATURE SELECTION

- Ridge Regression was applied to select the top 15 variables that contribute most to the target variable based on coefficient values (Figure 13)
- Cut off value for top variables is 0.1626726
- Top variable list (Figure 14)
- 16 variables after feature selection (Figure 15)

```r
# Ridge Regression

# Create a matrix from the predictor variables
x <- model.matrix(as.factor(df1$poisonous) ~ ., df1)[,-1] # We exclude the intercept column using [,-1]

# Create a vector from the target variable
y <- df1$poisonous

# Compute Ridge Regression
set.seed(234)  # for reproducibility
cv.ridge <- cv.glmnet(x, y, alpha = 0, nfolds = 10)

# Output the lambda that gives the minimum mean cross-validated error
lambda_min <- cv.ridge$lambda.min
lambda_min

# Fit the final model on the data using lambda.lasso
model.ridge <- glmnet(x, y, alpha = 0, lambda = lambda_min)

# Print coefficients of the Ridge Regression model
print(coef(model.ridge))

# Extract the coefficients
ridge_coefs <- as.data.frame(as.matrix(coef(model.ridge)))

# Make the row names into a column for variable names
ridge_coefs$variable <- rownames(ridge_coefs)

# Rename the V1 column to coefficient
colnames(ridge_coefs)[colnames(ridge_coefs)=="s0"] <- "coefficient"

# Remove the first row (which is the intercept term)
ridge_coefs <- ridge_coefs[-1, ]

# Create a new column with the absolute value of the coefficients
ridge_coefs$abs_coefficient <- abs(ridge_coefs$coefficient)
```

```r
# Order the coefficients by their absolute values, in descending order
ridge_coefs <- ridge_coefs[order(-ridge_coefs$abs_coefficient), ]

# Print the top variables
top_variables <- head(ridge_coefs, 15)
print(top_variables)

# Get the smallest absolute coefficient among the top 30 variables
cutoff <- min(top_variables$abs_coefficient)
cat("Cutoff for top variables: ", cutoff, "\n")

# Get the variable names of the top variables
top_var_names <- as.character(top_variables$variable)
top_var_names_clean <- gsub(".*`(.*)`", "\\1", top_variables$variable)
df3 <- cbind(df1[, top_var_names_clean], df1$poisonous) # Create df2 with only the top variables
```

Figure 13 Ridge Regression for Feature Selection

	coefficient	variable	abs_coefficient
`stalk_color_above_ring_df[[var]]yellow`	0.7974901	`stalk_color_above_ring_df[[var]]yellow`	0.7974901
`gill_color_df[[var]]green`	0.5630947	`gill_color_df[[var]]green`	0.5630947
`odor_df[[var]]none`	−0.4833235	`odor_df[[var]]none`	0.4833235
`stalk_color_below_ring_df[[var]]yellow`	0.4256052	`stalk_color_below_ring_df[[var]]yellow`	0.4256052
`stalk_surface_above_ring_df[[var]]scaly`	−0.3835527	`stalk_surface_above_ring_df[[var]]scaly`	0.3835527
`odor_df[[var]]anise`	−0.3557441	`odor_df[[var]]anise`	0.3557441
`odor_df[[var]]almond`	−0.3557371	`odor_df[[var]]almond`	0.3557371
`odor_df[[var]]creosote`	0.3533332	`odor_df[[var]]creosote`	0.3533332
`odor_df[[var]]pungent`	0.3506031	`odor_df[[var]]pungent`	0.3506031
`ring_type_df[[var]]flaring`	−0.3398938	`ring_type_df[[var]]flaring`	0.3398938
`stalk_root_df[[var]]rooted`	−0.3157500	`stalk_root_df[[var]]rooted`	0.3157500
`habitat_df[[var]]waste`	−0.2691463	`habitat_df[[var]]waste`	0.2691463
`stalk_root_df[[var]]club`	−0.2456299	`stalk_root_df[[var]]club`	0.2456299
`ring_type_df[[var]]pendant`	−0.1693763	`ring_type_df[[var]]pendant`	0.1693763
`habitat_df[[var]]urban`	0.1626726	`habitat_df[[var]]urban`	0.1626726

Figure 14 Top Variable Names

```{r}
str(df3)
```

```
'data.frame':   8124 obs. of  16 variables:
 $ stalk_color_above_ring_df[[var]]yellow : num  0 0 0 0 0 0 0 0 0 0 ...
 $ gill_color_df[[var]]green               : num  0 0 0 0 0 0 0 0 0 0 ...
 $ odor_df[[var]]none                      : num  0 0 0 0 1 0 0 0 0 0 ...
 $ stalk_color_below_ring_df[[var]]yellow : num  0 0 0 0 0 0 0 0 0 0 ...
 $ stalk_surface_above_ring_df[[var]]scaly: num  0 0 0 0 0 0 0 0 0 0 ...
 $ odor_df[[var]]anise                     : num  0 0 1 0 0 0 0 1 0 0 ...
 $ odor_df[[var]]almond                    : num  0 1 0 0 0 1 1 0 0 1 ...
 $ odor_df[[var]]creosote                  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ odor_df[[var]]pungent                   : num  1 0 0 1 0 0 0 0 1 0 ...
 $ ring_type_df[[var]]flaring              : num  0 0 0 0 0 0 0 0 0 0 ...
 $ stalk_root_df[[var]]rooted              : num  0 0 0 0 0 0 0 0 0 0 ...
 $ habitat_df[[var]]waste                  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ stalk_root_df[[var]]club                : num  0 1 1 0 0 1 1 1 0 1 ...
 $ ring_type_df[[var]]pendant              : num  1 1 1 1 0 1 1 1 1 1 ...
 $ habitat_df[[var]]urban                  : num  1 0 0 1 0 0 0 0 0 0 ...
 $ df1$poisonous                           : int  1 0 0 1 0 0 0 0 1 0 ...
```

Figure 15 Data Structure Post Feature Selection

VISUALIZATION WITH SELECTED FEATURES

CORRELATION AMONG SELECTED FEATURES
- Correlation Among Features (Figure 16) are plotted using cor() and we can identify positive relationships (red) and negative relationships. Features are chosen if they have more color (either red or green) while white color indicate the features don't have strong relationships associated with poisonous mushrooms
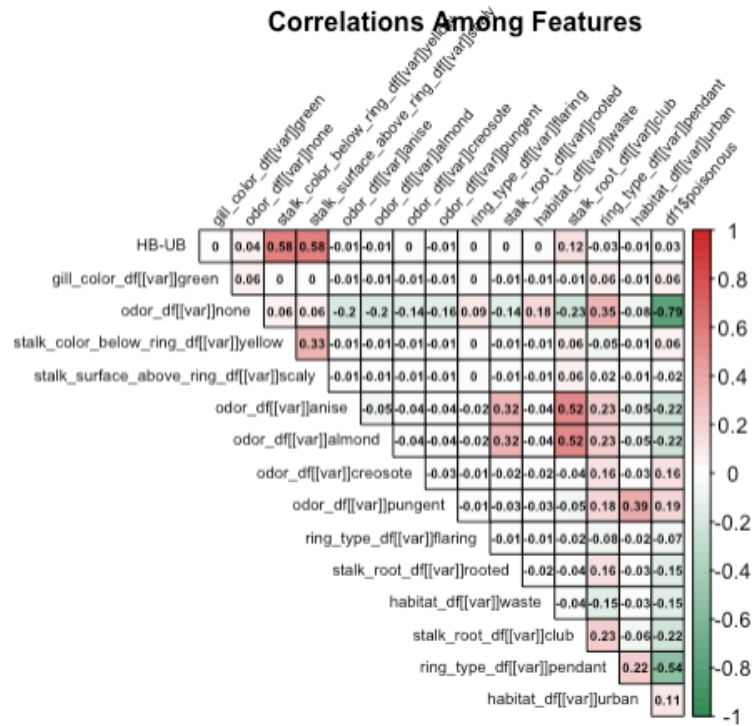

Figure 16 Correlation between the top variables and the target

HISTOGRAMS AMONG SELECTED FEATURES
- Histograms Among Features (Figure 17) are plotted to understand the trends in the top 15 features
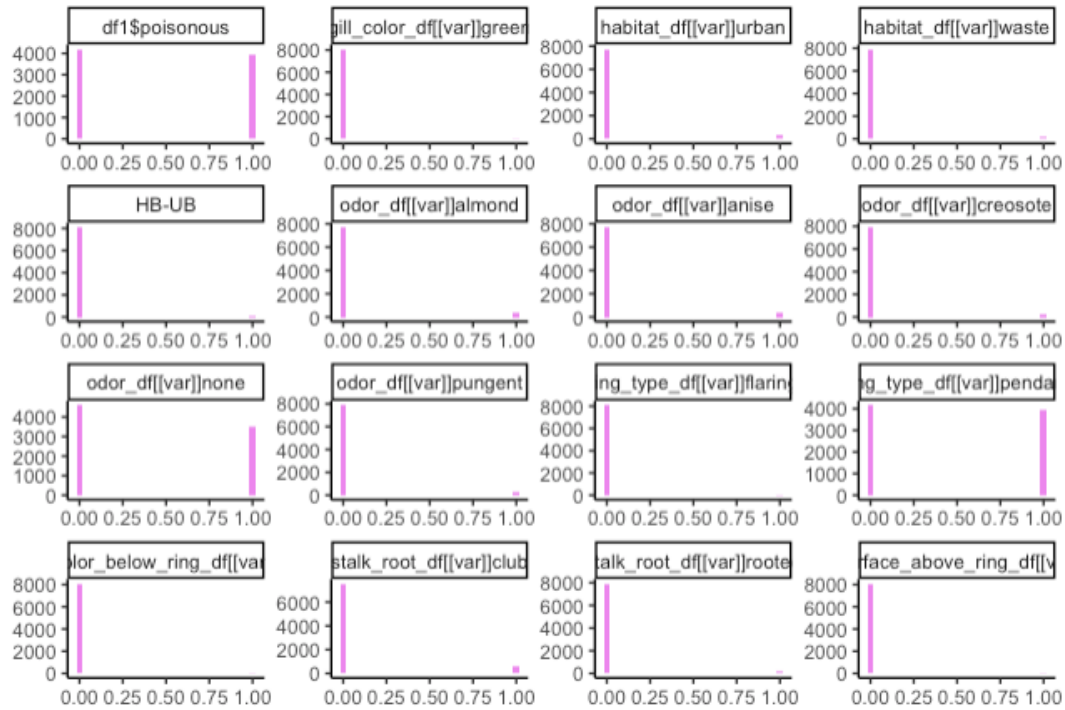
Figure 17 Top Variables Histograms

- Snapshot of the selected final dataset (Figure 18)

```{r}
summary(df3)
```

```
       HB-UB            gill_color_df[[var]]green odor_df[[var]]none stalk_color_below_ring_df[[var]]yello
 Min.   :0.0000000   Min.   :0.000000         Min.   :0.0000      Min.   :0.000000
 1st Qu.:0.0000000   1st Qu.:0.000000         1st Qu.:0.0000      1st Qu.:0.000000
 Median :0.0000000   Median :0.000000         Median :0.0000      Median :0.000000
 Mean   :0.0009847   Mean   :0.002954         Mean   :0.4343      Mean   :0.002954
 3rd Qu.:0.0000000   3rd Qu.:0.000000         3rd Qu.:1.0000      3rd Qu.:0.000000
 Max.   :1.0000000   Max.   :1.000000         Max.   :1.0000      Max.   :1.000000
 stalk_surface_above_ring_df[[var]]scaly odor_df[[var]]anise odor_df[[var]]almond odor_df[[var]]creosot
 Min.   :0.000000                         Min.   :0.00000    Min.   :0.00000     Min.   :0.00000
 1st Qu.:0.000000                         1st Qu.:0.00000    1st Qu.:0.00000     1st Qu.:0.00000
 Median :0.000000                         Median :0.00000    Median :0.00000     Median :0.00000
 Mean   :0.002954                         Mean   :0.04924    Mean   :0.04924     Mean   :0.02363
 3rd Qu.:0.000000                         3rd Qu.:0.00000    3rd Qu.:0.00000     3rd Qu.:0.00000
 Max.   :1.000000                         Max.   :1.00000    Max.   :1.00000     Max.   :1.00000
 odor_df[[var]]pungent ring_type_df[[var]]flaring stalk_root_df[[var]]rooted habitat_df[[var]]waste
 Min.   :0.00000       Min.   :0.000000           Min.   :0.00000           Min.   :0.00000
 1st Qu.:0.00000       1st Qu.:0.000000           1st Qu.:0.00000           1st Qu.:0.00000
 Median :0.00000       Median :0.000000           Median :0.00000           Median :0.00000
 Mean   :0.03151       Mean   :0.005908           Mean   :0.02363           Mean   :0.02363
 3rd Qu.:0.00000       3rd Qu.:0.000000           3rd Qu.:0.00000           3rd Qu.:0.00000
 Max.   :1.00000       Max.   :1.000000           Max.   :1.00000           Max.   :1.00000
 stalk_root_df[[var]]club ring_type_df[[var]]pendant habitat_df[[var]]urban poisonous
 Min.   :0.00000          Min.   :0.0000             Min.   :0.0000         0:4208
 1st Qu.:0.00000          1st Qu.:0.0000             1st Qu.:0.0000         1:3916
 Median :0.00000          Median :0.0000             Median :0.0000
 Mean   :0.06844          Mean   :0.4884             Mean   :0.0453
 3rd Qu.:0.00000          3rd Qu.:1.0000             3rd Qu.:0.0000
 Max.   :1.00000          Max.   :1.0000             Max.   :1.0000
```

```{r}
str(df3)
```

```
'data.frame':   8124 obs. of  16 variables:
 $ HB-UB                                  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ gill_color_df[[var]]green              : num  0 0 0 0 0 0 0 0 0 0 ...
 $ odor_df[[var]]none                     : num  0 0 0 0 1 0 0 0 0 0 ...
 $ stalk_color_below_ring_df[[var]]yellow : num  0 0 0 0 0 0 0 0 0 0 ...
 $ stalk_surface_above_ring_df[[var]]scaly: num  0 0 0 0 0 0 0 0 0 0 ...
 $ odor_df[[var]]anise                    : num  0 0 1 0 0 0 0 1 0 0 ...
 $ odor_df[[var]]almond                   : num  0 1 0 0 0 1 1 0 0 1 ...
 $ odor_df[[var]]creosote                 : num  0 0 0 0 0 0 0 0 0 0 ...
 $ odor_df[[var]]pungent                  : num  1 0 0 1 0 0 0 0 1 0 ...
 $ ring_type_df[[var]]flaring             : num  0 0 0 0 0 0 0 0 0 0 ...
 $ stalk_root_df[[var]]rooted             : num  0 0 0 0 0 0 0 0 0 0 ...
 $ habitat_df[[var]]waste                 : num  0 0 0 0 0 0 0 0 0 0 ...
 $ stalk_root_df[[var]]club               : num  0 1 1 0 0 1 1 1 0 1 ...
 $ ring_type_df[[var]]pendant             : num  1 1 1 1 0 1 1 1 1 1 ...
 $ habitat_df[[var]]urban                 : num  1 0 0 1 0 0 0 0 0 0 ...
 $ poisonous                              : Factor w/ 2 levels "0","1": 2 1 1 2 1 1 1 1 2 1 ...
```

Figure 18 Selected Data Summary and Structure

RESULT ANALYSIS

DATA PARTITION

We partition the data using the CARET package in R in training and test sets. The partition index is set at "0.80," which refers to the training set of 80% of the whole dataset and the test set of 20% (Figure 19).

```
set.seed(123)
df2$dv_int <- as.factor(df2$dv_int)
trainingIndex <- createDataPartition(df2$dv_int, p=0.8, list = FALSE)
df_train <- df2[trainingIndex, ]
df_test <- df2[- trainingIndex,]
```

Figure 19 Data partition

PERFORM LM
According to LM, Gill_color_green, odor_none, stalk_color_below_ring_yellow, odor_anise and odor_almond are *** significant variables. Habitat_waste is ** significant variable (Figure 20).

```
Call:
lm(formula = poisonous ~ ., data = df3_norm)

Residuals:
     Min       1Q    Median       3Q       Max
-0.04768 -0.04768 -0.00004 -0.00004  1.96282

Coefficients:
                                              Estimate Std. Error  t value Pr(>|t|)
(Intercept)                                  2.799e-14  2.068e-03    0.000   1.0000
`HB-UB`                                      1.496e-03  2.976e-03    0.503   0.6152
`gill_color_df[[var]]green`                  1.060e-01  2.075e-03   51.088   <2e-16 ***
`odor_df[[var]]none`                        -9.730e-01  3.149e-03 -309.000   <2e-16 ***
`stalk_color_below_ring_df[[var]]yellow`     1.065e-01  2.549e-03   41.792   <2e-16 ***
`stalk_surface_above_ring_df[[var]]scaly`   -2.588e-03  2.538e-03   -1.020   0.3079
`odor_df[[var]]anise`                       -4.350e-01  4.088e-03 -106.420   <2e-16 ***
`odor_df[[var]]almond`                      -4.350e-01  4.088e-03 -106.420   <2e-16 ***
`odor_df[[var]]creosote`                    -1.422e-03  2.303e-03   -0.618   0.5369
`odor_df[[var]]pungent`                      6.686e-05  2.411e-03    0.028   0.9779
`ring_type_df[[var]]flaring`                -2.939e-03  2.110e-03   -1.393   0.1636
`stalk_root_df[[var]]rooted`                -1.425e-06  3.222e-03    0.000   0.9996
`habitat_df[[var]]waste`                    -5.826e-03  2.227e-03   -2.616   0.0089 **
`stalk_root_df[[var]]club`                  -2.812e-06  4.362e-03   -0.001   0.9995
`ring_type_df[[var]]pendant`                 4.662e-03  3.300e-03    1.413   0.1577
`habitat_df[[var]]urban`                    -4.054e-03  2.372e-03   -1.709   0.0875 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1864 on 8108 degrees of freedom
Multiple R-squared:  0.9653,    Adjusted R-squared:  0.9652
F-statistic: 1.504e+04 on 15 and 8108 DF,  p-value: < 2.2e-16
```

Figure 20 LM

PERFORM ANOVA

According to ANOVA, habitat-urban, gill_color_green, odor_none, stalk_color_below_ring_yellow, odor_anise and odor_almond are *** significant variables. Habitat_waste is ** significant variable (Figure 21). The results are the same with LM model except for the habitat urban variable.

```
                                          Df Sum Sq Mean Sq   F value  Pr(>F)
`HB-UB`                                    1      9       9 2.476e+02 < 2e-16 ***
`gill_color_df[[var]]green`                1     26      26 7.457e+02 < 2e-16 ***
`odor_df[[var]]none`                       1   5099    5099 1.467e+05 < 2e-16 ***
`stalk_color_below_ring_df[[var]]yellow`   1     61      61 1.760e+03 < 2e-16 ***
`stalk_surface_above_ring_df[[var]]scaly`  1      0       0 7.930e-01 0.37325
`odor_df[[var]]anise`                      1   1197    1197 3.444e+04 < 2e-16 ***
`odor_df[[var]]almond`                     1   1449    1449 4.170e+04 < 2e-16 ***
`odor_df[[var]]creosote`                   1      0       0 0.000e+00 1.00000
`odor_df[[var]]pungent`                    1      0       0 0.000e+00 1.00000
`ring_type_df[[var]]flaring`               1      0       0 2.423e+00 0.11958
`stalk_root_df[[var]]rooted`               1      0       0 0.000e+00 1.00000
`habitat_df[[var]]waste`                   1      0       0 1.041e+01 0.00126 **
`stalk_root_df[[var]]club`                 1      0       0 0.000e+00 1.00000
`ring_type_df[[var]]pendant`               1      0       0 8.720e-01 0.35052
`habitat_df[[var]]urban`                   1      0       0 2.921e+00 0.08749 .
Residuals                               8108    282       0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 21 ANOVA