Predicting Mushroom Edibility

Justin Farnsworth

7/2/2020

Summary

In 1981, The Audubon Society Field Guide to North American Mushrooms created a data set consisting of 23 different types of mushrooms. These mushrooms were categorized into two groups based on their edibility: edible or poisonous. Mushrooms that were not verifiably edible or not safe to eat were considered to be poisonous. In this project, we predicted whether a mushroom is safe to eat.

A brief exploration of the data was conducted, which found that odor was a good indication of a mushroom's edibility. Depending on the odor, the mushroom can be safe to eat, according to the data. For example, mushrooms with a foul, fishy, or spicy odor were not safe to eat. Meanwhile, mushrooms that smelled like almonds or anise were considered safe to eat. The only area of uncertainty here was when the mushrooms didn't have an odor.

Another feature that is useful for determining edibility is the gill color. Mushrooms with green and buff gill colors were poisonous while those with orange and red gill colors were edible. Similarly, the stalk color above the mushroom's ring (annulus) can help determine edibility as well, with buff, cinnamon, and yellow being unsafe to eat whereas gray, orange, and red were safe.

After splitting the data into a training set (80%) and a test set (20%), we used three models to predict mushroom edibility: logistic regeression, k-nearest neighbors, and a classification tree. We achieved an **accuracy of 100%** for each model! However, there were warnings presented when fitting the logistic regression model.

The data set can be accessed here: https://www.kaggle.com/uciml/mushroom-classification

Analysis

A copy of the data set can be downloaded from this project's GitHub repository. (https://github.com/farnswj1/Predicting_Mushroom_Edibility)

Cleaning Up the Data Set

```
# Required packages
if(!require(tidyverse)) install.packages("tidyverse", repos = "http://cran.us.r-project.org")
if(!require(caret)) install.packages("caret", repos = "http://cran.us.r-project.org")
if(!require(ggthemes)) install.packages("ggthemes", repos = "http://cran.us.r-project.org")


# Original Source: https://www.kaggle.com/uciml/mushroom-classification
```

```
temp <- tempfile()
download.file(
  "https://raw.github.com/farnswj1/Predicting_Mushroom_Edibility/master/mushrooms.csv",
  temp
)

# Load the data set remove the tempfile
data <- read.csv(temp)
rm(temp)
```

There are 8124 rows and 23 columns in the data set. When looking at the data set, the values for each column are unclear as each value is represented as one character.

```
# Show first 10 rows of the data set before cleanup
head(data, 10)
```

```
##     class cap.shape cap.surface cap.color bruises odor gill.attachment
## 1       p         x           s         n       t    p               f
## 2       e         x           s         y       t    a               f
## 3       e         b           s         w       t    l               f
## 4       p         x           y         w       t    p               f
## 5       e         x           s         g       f    n               f
## 6       e         x           y         y       t    a               f
## 7       e         b           s         w       t    a               f
## 8       e         b           y         w       t    l               f
## 9       p         x           y         w       t    p               f
## 10      e         b           s         y       t    a               f
##     gill.spacing gill.size gill.color stalk.shape stalk.root
## 1              c         n          k           e          e
## 2              c         b          k           e          c
## 3              c         b          n           e          c
## 4              c         n          n           e          e
## 5              w         b          k           t          e
## 6              c         b          n           e          c
## 7              c         b          g           e          c
## 8              c         b          n           e          c
## 9              c         n          p           e          e
## 10             c         b          g           e          c
##     stalk.surface.above.ring stalk.surface.below.ring stalk.color.above.ring
## 1                          s                        s                      w
## 2                          s                        s                      w
## 3                          s                        s                      w
## 4                          s                        s                      w
## 5                          s                        s                      w
## 6                          s                        s                      w
## 7                          s                        s                      w
## 8                          s                        s                      w
## 9                          s                        s                      w
## 10                         s                        s                      w
##     stalk.color.below.ring veil.type veil.color ring.number ring.type
## 1                        w         p          w           o         p
## 2                        w         p          w           o         p
## 3                        w         p          w           o         p
```

```
## 4                          w         p         w         o         p
## 5                          w         p         w         o         e
## 6                          w         p         w         o         p
## 7                          w         p         w         o         p
## 8                          w         p         w         o         p
## 9                          w         p         w         o         p
## 10                         w         p         w         o         p
##     spore.print.color population habitat
## 1                  k          s       u
## 2                  n          n       g
## 3                  n          n       m
## 4                  k          s       u
## 5                  n          a       g
## 6                  k          n       g
## 7                  k          n       m
## 8                  n          s       m
## 9                  k          v       g
## 10                 k          s       m
```

We can also see that there appears to be no missing values in the data set.

```
# Check for null values
any(is.na(data))
```

```
## [1] FALSE
```

Fortunately, the link to the data set on Kaggle (see https://www.kaggle.com/uciml/mushroom-classification) defines the values for each character. We cleaned up the data set so that the values are replaced with the proper descriptors.

After cleaning up the data set, we have the following:

```
# Show first 10 rows of the data set after cleanup
head(data, 10)
```

```
##     edibility cap.shape cap.surface cap.color bruises    odor gill.attachment
## 1  Poisonous    Convex      Smooth     Brown    TRUE Pungent            Free
## 2     Edible    Convex      Smooth    Yellow    TRUE  Almond            Free
## 3     Edible      Bell      Smooth     White    TRUE   Anise            Free
## 4  Poisonous    Convex       Scaly     White    TRUE Pungent            Free
## 5     Edible    Convex      Smooth      Gray   FALSE    None            Free
## 6     Edible    Convex       Scaly    Yellow    TRUE  Almond            Free
## 7     Edible      Bell      Smooth     White    TRUE  Almond            Free
## 8     Edible      Bell       Scaly     White    TRUE   Anise            Free
## 9  Poisonous    Convex       Scaly     White    TRUE Pungent            Free
## 10    Edible      Bell      Smooth    Yellow    TRUE  Almond            Free
##     gill.spacing gill.size gill.color stalk.shape stalk.root
## 1          Close    Narrow      Black   Enlarging      Equal
## 2          Close     Broad      Black   Enlarging       Club
## 3          Close     Broad      Brown   Enlarging       Club
## 4          Close    Narrow      Brown   Enlarging      Equal
## 5        Crowded     Broad      Black    Tapering      Equal
## 6          Close     Broad      Brown   Enlarging       Club
```

```
## 7           Close       Broad     Gray   Enlarging       Club
## 8           Close       Broad    Brown   Enlarging       Club
## 9           Close      Narrow     Pink   Enlarging      Equal
## 10          Close       Broad     Gray   Enlarging       Club
##     stalk.surface.above.ring stalk.surface.below.ring stalk.color.above.ring
## 1                     Smooth                   Smooth                  White
## 2                     Smooth                   Smooth                  White
## 3                     Smooth                   Smooth                  White
## 4                     Smooth                   Smooth                  White
## 5                     Smooth                   Smooth                  White
## 6                     Smooth                   Smooth                  White
## 7                     Smooth                   Smooth                  White
## 8                     Smooth                   Smooth                  White
## 9                     Smooth                   Smooth                  White
## 10                    Smooth                   Smooth                  White
##     stalk.color.below.ring veil.color ring.number  ring.type spore.print.color
## 1                    White      White        One    Pendant             Black
## 2                    White      White        One    Pendant             Brown
## 3                    White      White        One    Pendant             Brown
## 4                    White      White        One    Pendant             Black
## 5                    White      White        One Evanescent             Brown
## 6                    White      White        One    Pendant             Black
## 7                    White      White        One    Pendant             Black
## 8                    White      White        One    Pendant             Brown
## 9                    White      White        One    Pendant             Black
## 10                   White      White        One    Pendant             Black
##     population habitat
## 1    Scattered   Urban
## 2     Numerous Grasses
## 3     Numerous Meadows
## 4    Scattered   Urban
## 5     Abundant Grasses
## 6     Numerous Grasses
## 7     Numerous Meadows
## 8    Scattered Meadows
## 9      Several Grasses
## 10   Scattered Meadows
```

Note that the `veil.type` column was removed as it contained only one unique value. It would not be useful for predicting mushroom edibility later.

Exploring the Data Set

The following R code was used to generate summary data of each column. Each column's summary contains the total number of mushrooms that have a specific value as well as the percentage of edible mushrooms with that column value.

```r
# Create a list with the totals and percentage of edible mushrooms for each group in the column
tables <- map(colnames(data), function(column){
  # Create a table of totals and percentage of edible mushrooms for each group
  table <- data %>%
    group_by(data[, column]) %>%
    summarize(total = n(), percentage = mean(edibility == "Edible") * 100)
```

```
  # Clean up the column names
  setNames(table, c(column, "total", "percentage"))
})

# Display the tables
tables
```

```
## [[1]]
## # A tibble: 2 x 3
##   edibility total percentage
##   <fct>     <int>      <dbl>
## 1 Edible     4208        100
## 2 Poisonous  3916          0
##
## [[2]]
## # A tibble: 6 x 3
##   cap.shape total percentage
##   <fct>     <int>      <dbl>
## 1 Bell        452       89.4
## 2 Conical       4          0
## 3 Convex     3656       53.3
## 4 Flat       3152       50.6
## 5 Knobbed     828       27.5
## 6 Sunken       32        100
##
## [[3]]
## # A tibble: 4 x 3
##   cap.surface total percentage
##   <fct>       <int>      <dbl>
## 1 Fibrous      2320       67.2
## 2 Grooves         4          0
## 3 Scaly        3244       46.4
## 4 Smooth       2556       44.8
##
## [[4]]
## # A tibble: 10 x 3
##    cap.color total percentage
##    <fct>     <int>      <dbl>
##  1 Brown      2284       55.3
##  2 Buff        168       28.6
##  3 Cinnamon     44       72.7
##  4 Gray       1840       56.1
##  5 Green        16        100
##  6 Pink        144       38.9
##  7 Purple       16        100
##  8 Red        1500       41.6
##  9 White      1040       69.2
## 10 Yellow     1072       37.3
##
## [[5]]
## # A tibble: 2 x 3
##   bruises total percentage
##   <fct>   <int>      <dbl>
```

```
## 1 FALSE      4748       30.7
## 2 TRUE       3376       81.5
##
## [[6]]
## # A tibble: 9 x 3
##   odor      total percentage
##   <fct>     <int>      <dbl>
## 1 Almond      400        100
## 2 Anise       400        100
## 3 Creosote    192          0
## 4 Fishy       576          0
## 5 Foul       2160          0
## 6 Musty        36          0
## 7 None       3528       96.6
## 8 Pungent     256          0
## 9 Spicy       576          0
##
## [[7]]
## # A tibble: 2 x 3
##   gill.attachment total percentage
##   <fct>           <int>      <dbl>
## 1 Attached          210       91.4
## 2 Free             7914       50.7
##
## [[8]]
## # A tibble: 2 x 3
##   gill.spacing total percentage
##   <fct>        <int>      <dbl>
## 1 Close         6812       44.2
## 2 Crowded       1312       91.5
##
## [[9]]
## # A tibble: 2 x 3
##   gill.size total percentage
##   <fct>     <int>      <dbl>
## 1 Broad      5612       69.9
## 2 Narrow     2512       11.5
##
## [[10]]
## # A tibble: 12 x 3
##    gill.color total percentage
##    <fct>      <int>      <dbl>
##  1 Black        408       84.3
##  2 Brown       1048       89.3
##  3 Buff        1728          0
##  4 Chocolate    732       27.9
##  5 Gray         752       33.0
##  6 Green         24          0
##  7 Orange        64        100
##  8 Pink        1492       57.1
##  9 Purple       492       90.2
## 10 Red           96        100
## 11 White       1202       79.5
## 12 Yellow        86       74.4
```

```
##
## [[11]]
## # A tibble: 2 x 3
##   stalk.shape total percentage
##   <fct>       <int>      <dbl>
## 1 Enlarging    3516       46.0
## 2 Tapering     4608       56.2
##
## [[12]]
## # A tibble: 5 x 3
##   stalk.root total percentage
##   <fct>      <int>      <dbl>
## 1 Bulbous     3776       50.8
## 2 Club         556       92.1
## 3 Equal       1120       77.1
## 4 Missing     2480       29.0
## 5 Rooted       192      100
##
## [[13]]
## # A tibble: 4 x 3
##   stalk.surface.above.ring total percentage
##   <fct>                    <int>      <dbl>
## 1 Fibrous                    552       73.9
## 2 Scaly                       24       66.7
## 3 Silky                     2372        6.07
## 4 Smooth                    5176       70.3
##
## [[14]]
## # A tibble: 4 x 3
##   stalk.surface.below.ring total percentage
##   <fct>                    <int>      <dbl>
## 1 Fibrous                    600       76
## 2 Scaly                      284       73.2
## 3 Silky                     2304        6.25
## 4 Smooth                    4936       68.9
##
## [[15]]
## # A tibble: 9 x 3
##   stalk.color.above.ring total percentage
##   <fct>                  <int>      <dbl>
## 1 Brown                    448        3.57
## 2 Buff                     432        0
## 3 Cinnamon                  36        0
## 4 Gray                     576      100
## 5 Orange                   192      100
## 6 Pink                    1872       30.8
## 7 Red                       96      100
## 8 White                   4464       61.6
## 9 Yellow                     8        0
##
## [[16]]
## # A tibble: 9 x 3
##   stalk.color.below.ring total percentage
##   <fct>                  <int>      <dbl>
```

```
## 1 Brown                        512        12.5
## 2 Buff                         432          0
## 3 Cinnamon                      36          0
## 4 Gray                         576        100
## 5 Orange                       192        100
## 6 Pink                        1872         30.8
## 7 Red                           96        100
## 8 White                       4384         61.7
## 9 Yellow                        24          0
##
## [[17]]
## # A tibble: 4 x 3
##   veil.color total percentage
##   <fct>      <int>      <dbl>
## 1 Brown         96        100
## 2 Orange        96        100
## 3 White       7924         50.7
## 4 Yellow         8          0
##
## [[18]]
## # A tibble: 3 x 3
##   ring.number total percentage
##   <fct>       <int>      <dbl>
## 1 None           36          0
## 2 One          7488         49.1
## 3 Two           600         88
##
## [[19]]
## # A tibble: 5 x 3
##   ring.type    total percentage
##   <fct>        <int>      <dbl>
## 1 Evanescent    2776        36.3
## 2 Flaring         48        100
## 3 Large         1296          0
## 4 None            36          0
## 5 Pendant       3968         79.4
##
## [[20]]
## # A tibble: 9 x 3
##   spore.print.color total percentage
##   <fct>             <int>      <dbl>
## 1 Black              1872        88.0
## 2 Brown              1968        88.6
## 3 Buff                 48        100
## 4 Chocolate          1632         2.94
## 5 Green                72          0
## 6 Orange               48        100
## 7 Purple               48        100
## 8 White              2388         24.1
## 9 Yellow               48        100
##
## [[21]]
## # A tibble: 6 x 3
##   population total percentage
```

```
##   <fct>        <int>       <dbl>
## 1 Abundant       384       100
## 2 Clustered      340        84.7
## 3 Numerous       400       100
## 4 Scattered     1248        70.5
## 5 Several       4040        29.5
## 6 Solitary      1712        62.1
##
## [[22]]
## # A tibble: 7 x 3
##    habitat total percentage
##    <fct>   <int>      <dbl>
## 1 Grasses  2148        65.5
## 2 Leaves    832        28.8
## 3 Meadows   292        87.7
## 4 Paths    1144        11.9
## 5 Urban     368        26.1
## 6 Waste     192       100
## 7 Woods    3148        59.7
```
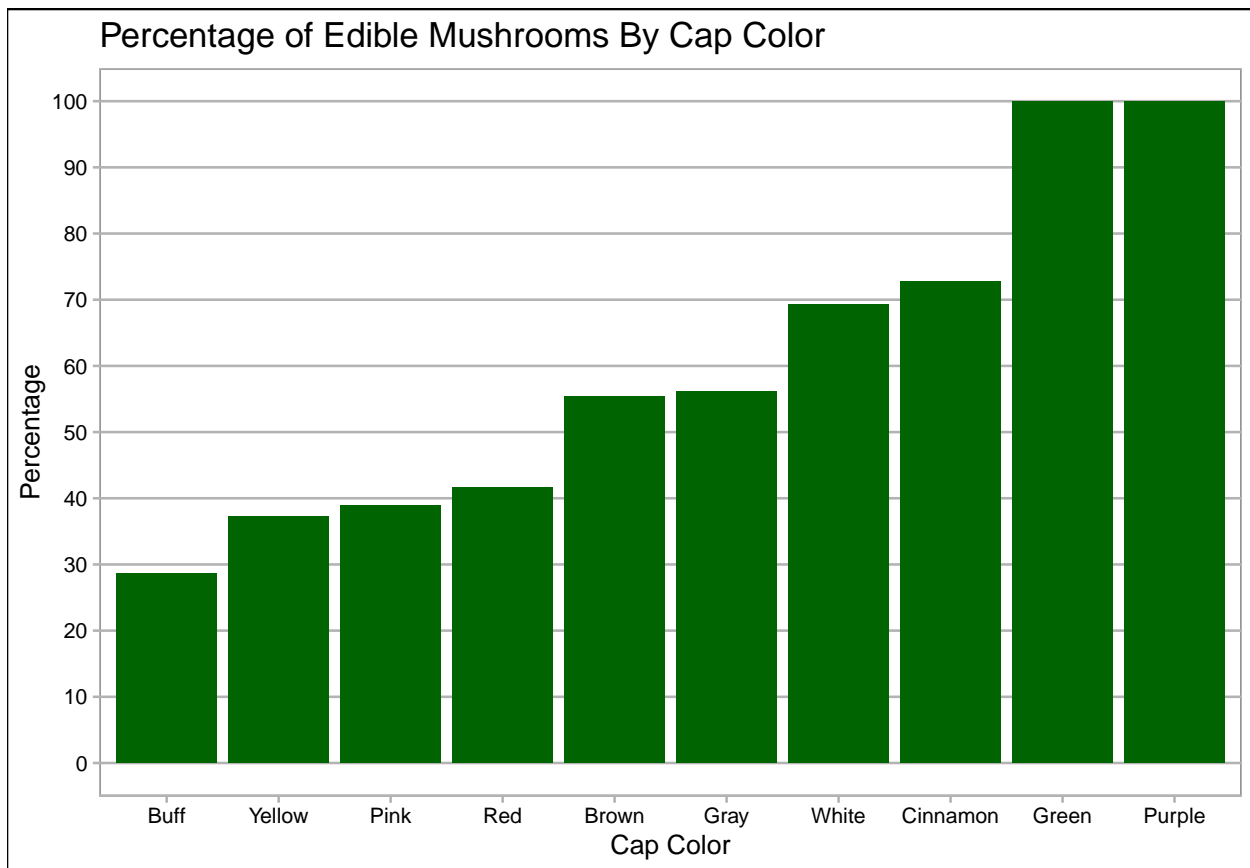
We can observe the following: Mushrooms with sunken cap shapes appear safe to eat, while conical cap shapes were not. However, the prevalence of both groups are low.
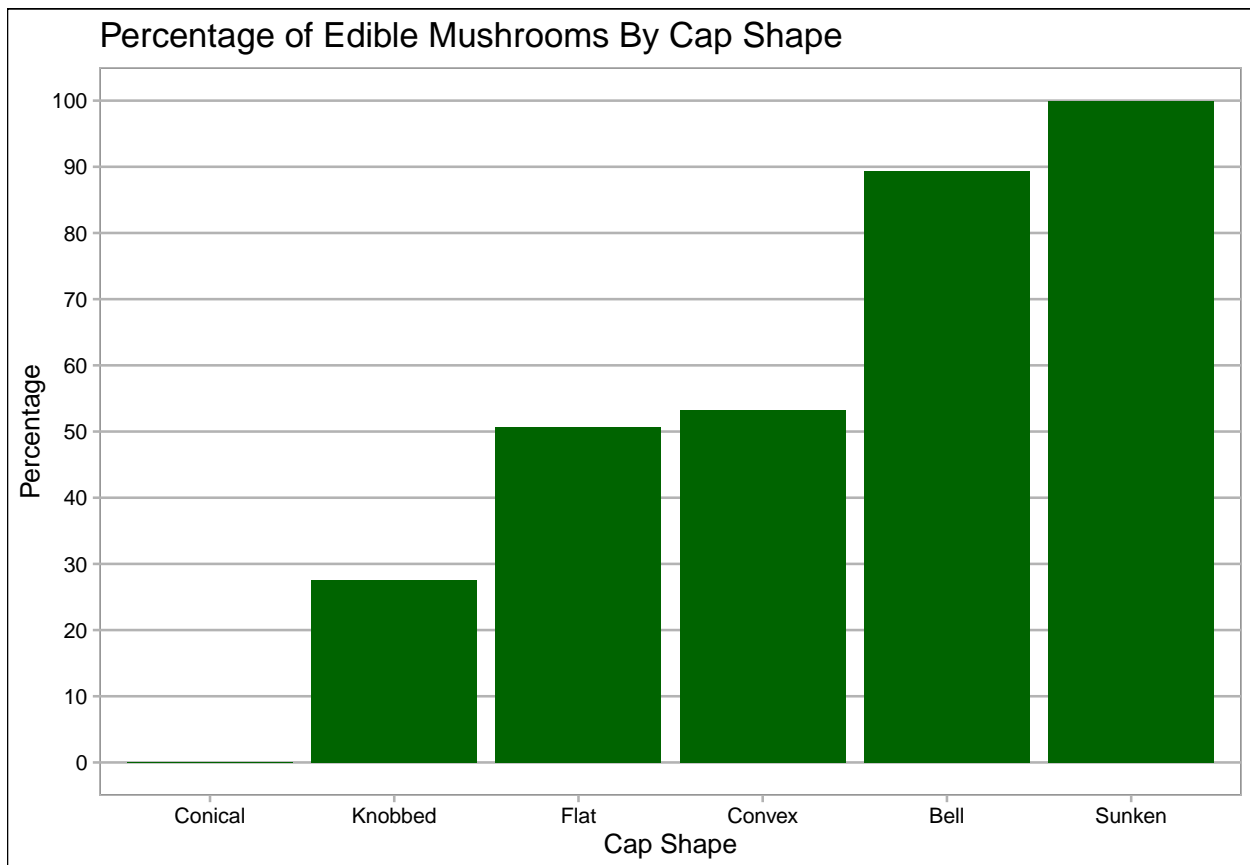
Mushrooms with grooved cap surfaces were considered not safe to eat. However, the prevalence of this groups is low.

Mushrooms with green and purple caps were safe to eat. However, the prevalence of both groups are low.

Percentage of Edible Mushrooms By Cap Color

A bar chart titled "Percentage of Edible Mushrooms By Cap Color" with the y-axis labeled "Percentage" (0 to 100) and the x-axis labeled "Cap Color".

Cap Color	Percentage
Buff	~29
Yellow	~37
Pink	~39
Red	~42
Brown	~55
Gray	~56
White	~69
Cinnamon	~73
Green	100
Purple	100

Mushrooms with bruises were more likely to be safe to eat than those without bruises.

Mushrooms that smelled like almond and anise were safe to eat. However, all other odors were not safe to eat. The exceptions were those that didn't have an odor, which were mostly safe to eat.

Mushrooms with attached gills were mostly safe to eat. Those that had free gills were more likely to be poisonous.

Percentage of Edible Mushrooms By Gill Attachment
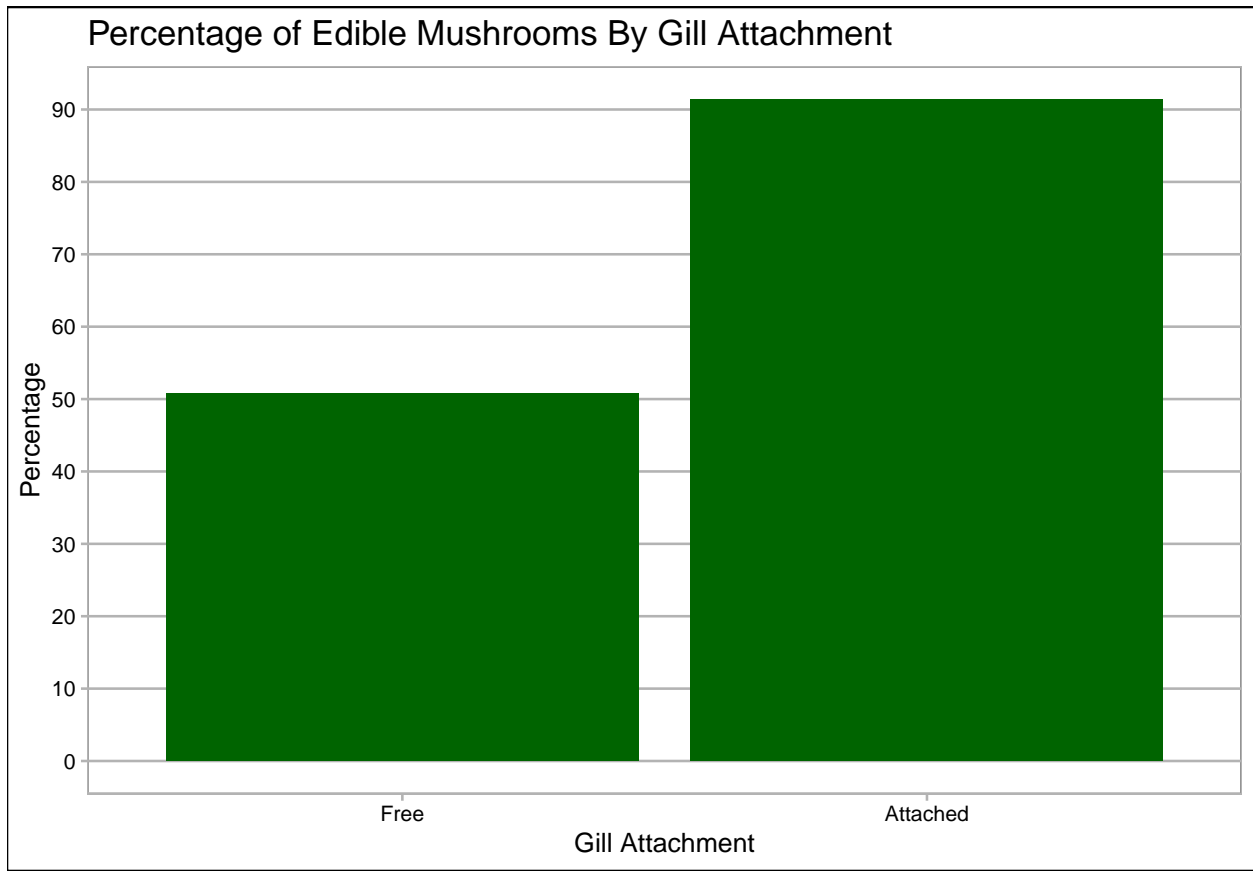
Mushrooms that were classified to have a crowded gill spacing were more likely to be edible than those that were classified as close.

Mushrooms that were classified to have a broad gill size were more likely to be edible than those that were considered to have a narrow gill size. The latter is generally unsafe to eat, according to the data.

Mushrooms with a buff or green gill color were not safe to eat. Meanwhile, those with orange or red gill colors were. Those with a purple gill color were mostly safe to eat.

Percentage of Edible Mushrooms By Gill Color
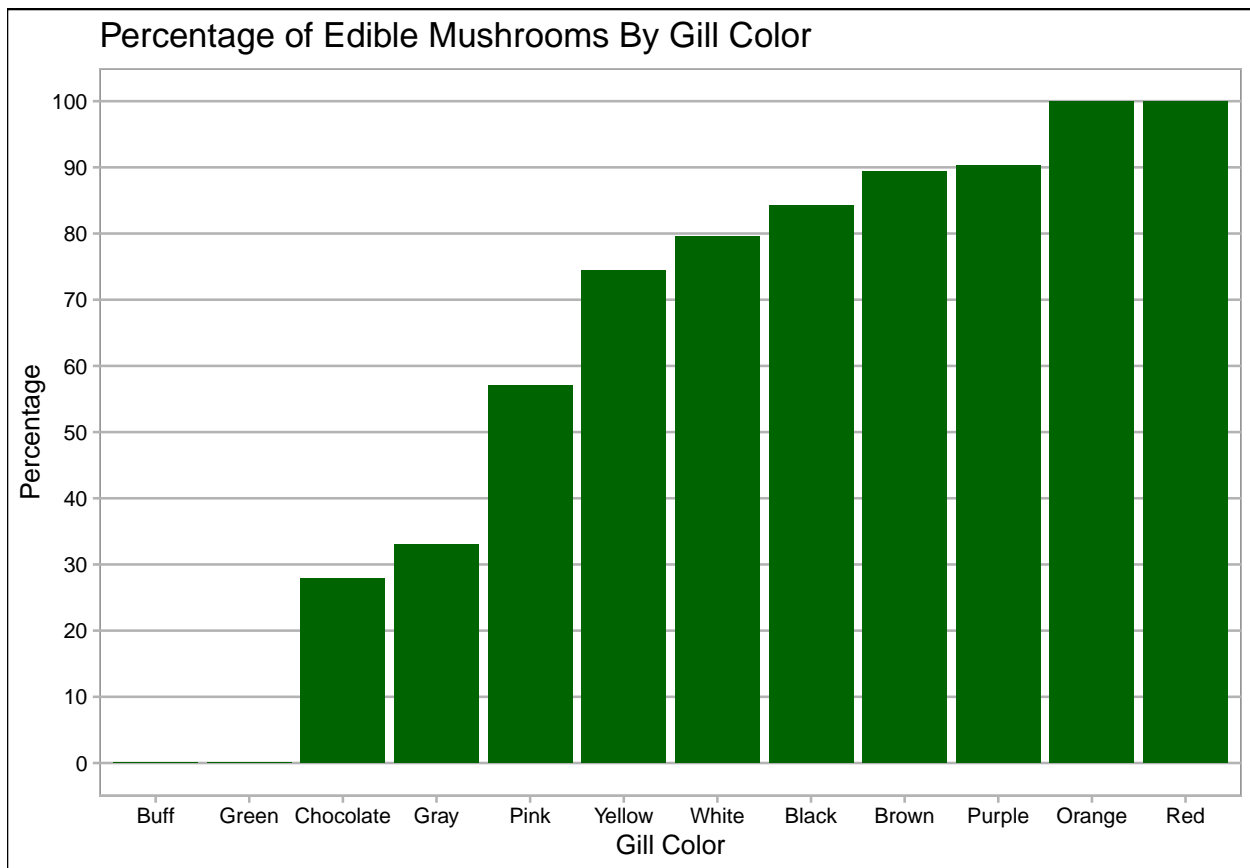
Mushrooms with a tapering stalk shape were more likely to be edible than those with an enlarging stalk shape, but not by much.

Mushrooms with a rooted stalk were safe to eat. Those classified as having a club or "equal" stalk root were mostly safe to eat.

Percentage of Edible Mushrooms By Stalk Root
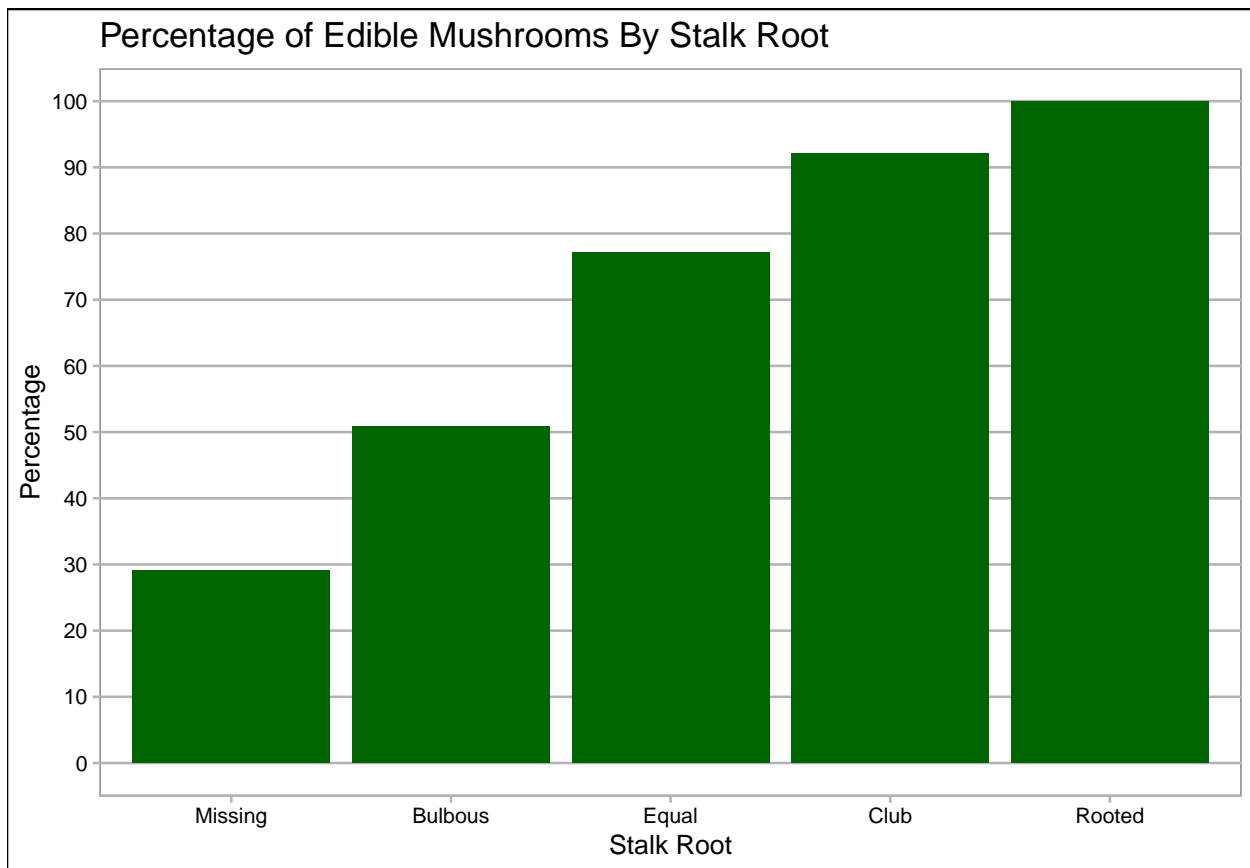
Mushrooms with a silky stalk surface are generally unsafe to eat. The others were more likely to be safe to eat, but not guaranteed.

Mushrooms with a buff, cinnamon, or yellow stalk color were not safe to eat. Those with gray, orange, and red stalk colors were safe to eat.

Mushrooms with a brown or orange veil color were safe to eat and those with a yellow veil color were not. However, the prevalence of these groups are low. Most mushrooms in the data set have a white veil color, while is approximately 50% likely to be safe to eat.

Mushrooms with no rings were not safe to eat. Nearly 50% of the mushrooms with one ring were safe to eat and 88% of those with two rings were safe to eat. It appears that more rings suggest a better chance of being edible.

Percentage of Edible Mushrooms By Number of Rings
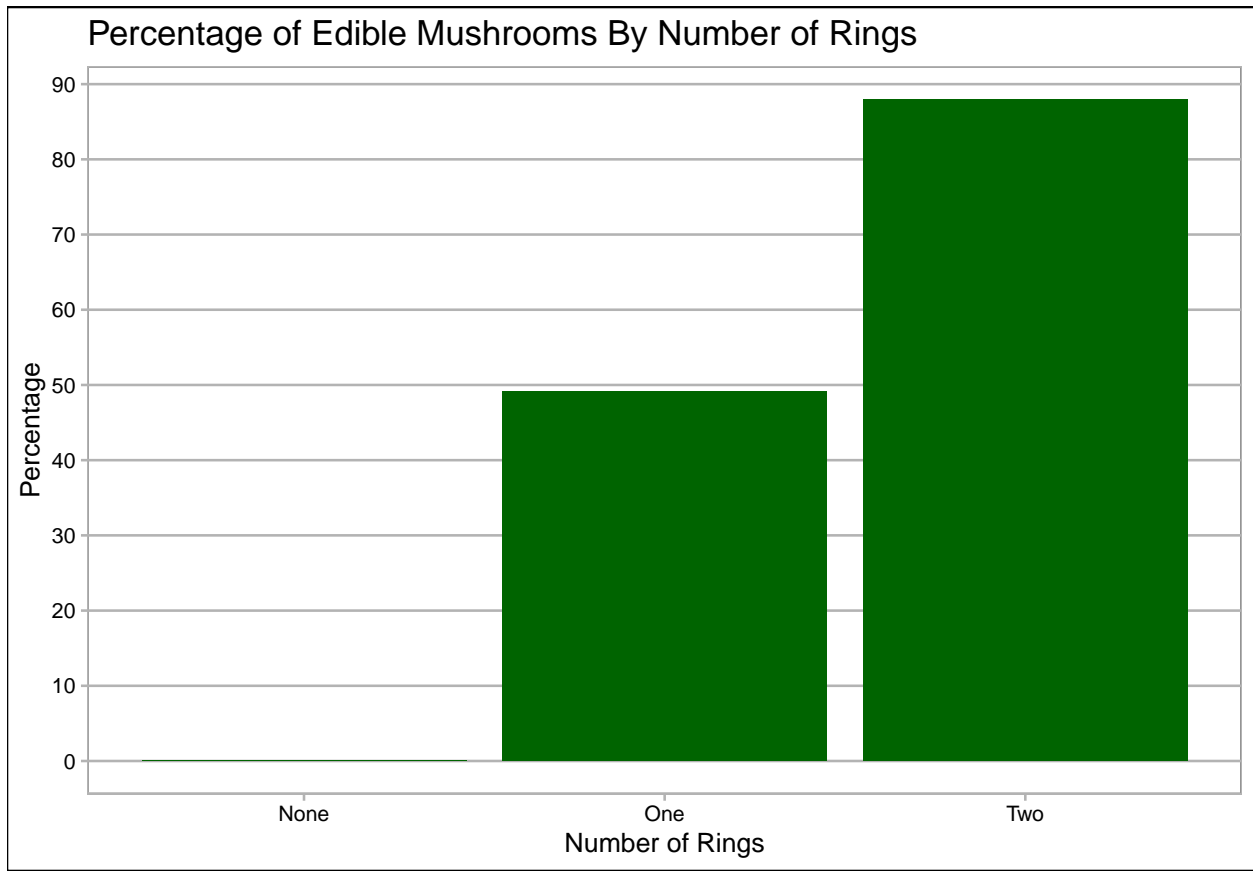
Mushrooms that were classified to have a large ring type were not safe to eat. Interestingly, there is a category called `None`, which is also unsafe to eat. Those with a flaring ring type were considered safe to eat, however this group has a low prevalence as well as `None`.

Percentage of Edible Mushrooms By Ring Type
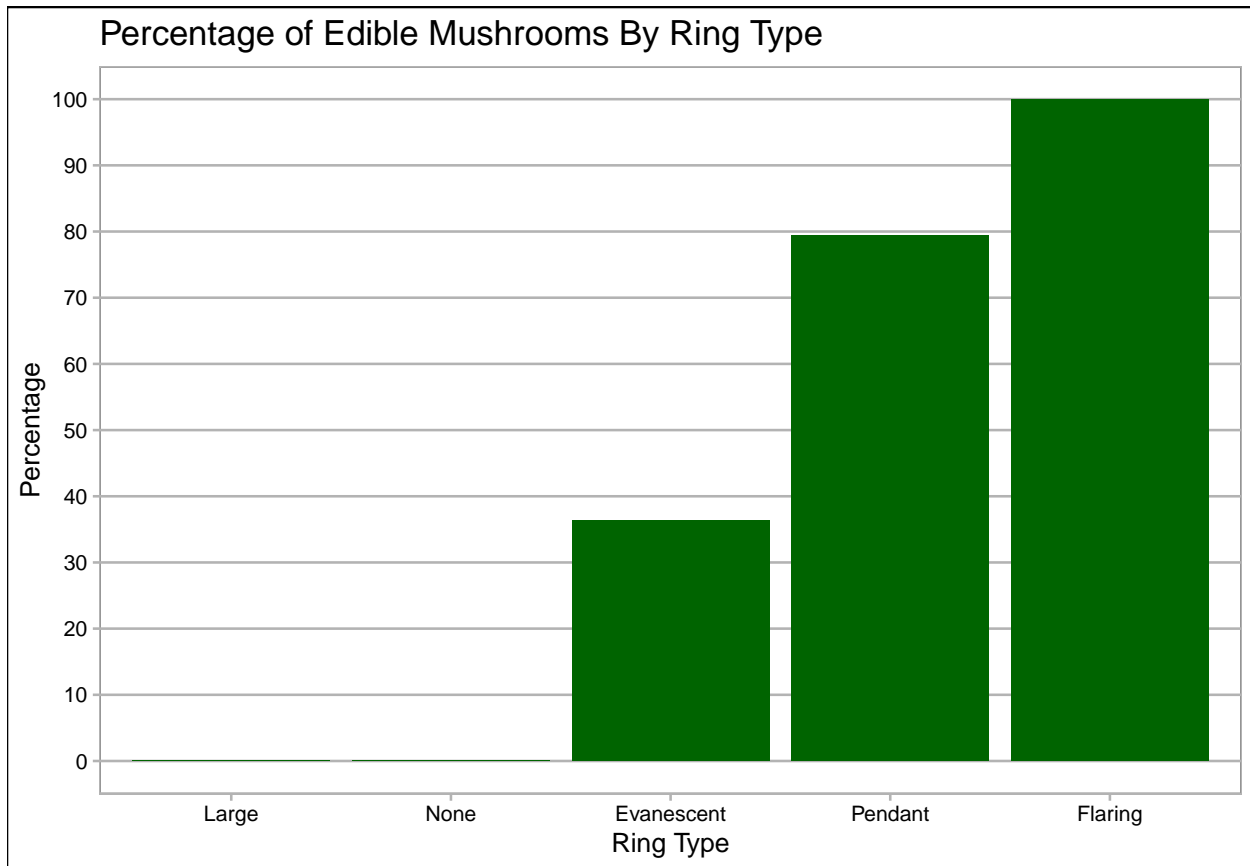
Mushrooms with a buff, orange, purple, and yellow spore print were safe to eat, while those with green spore prints were not safe to eat. All five of the mentioned categories have a low prevalence. Those with a chocolate spore print were generally unsafe to eat and those with a black or brown spore print were very likely to be safe to eat.

Mushrooms that were categorized to be adundant and numerous were safe to eat.

Percentage of Edible Mushrooms By Population Type
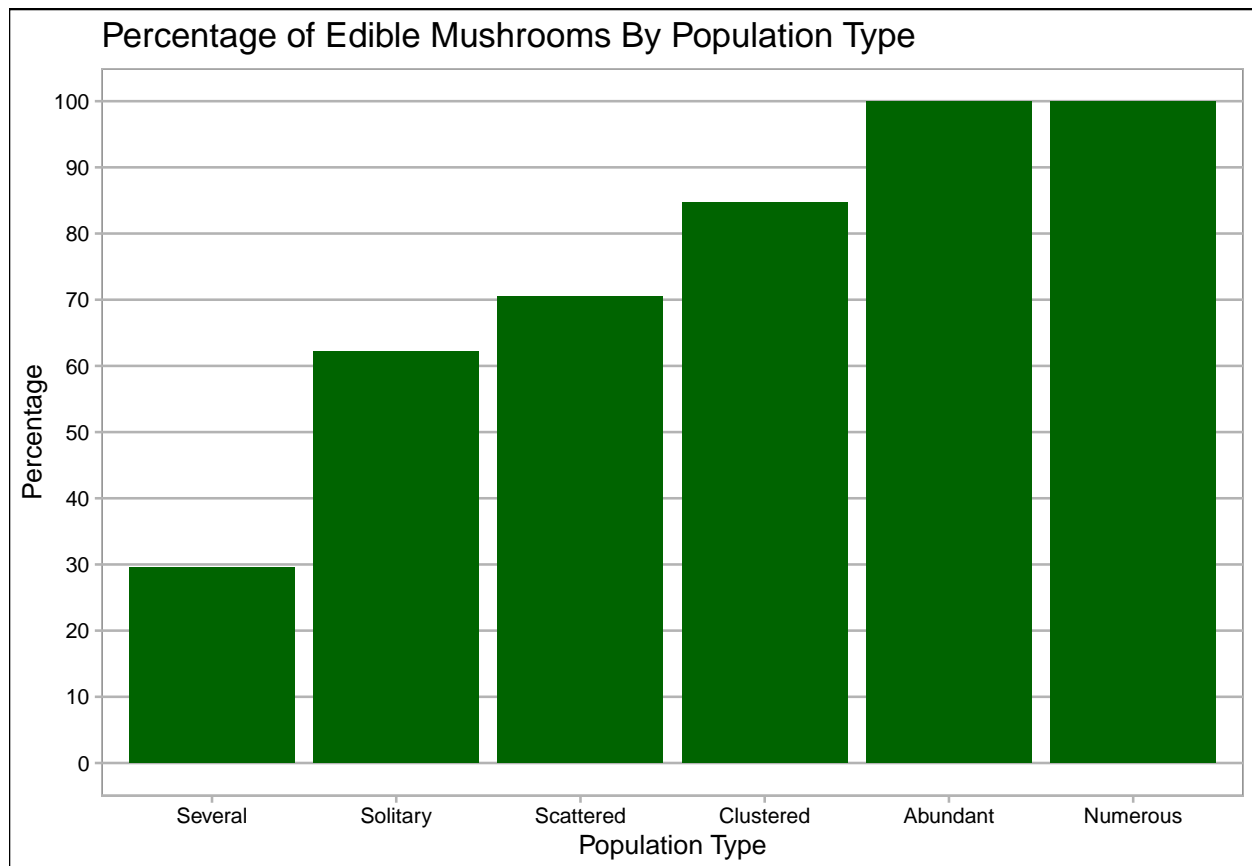
Mushrooms found on paths are generally not safe to eat. However, those found in waste were safe to eat. The most prevalent group here were those from the woods, in which nearly 60% of them were safe to eat.
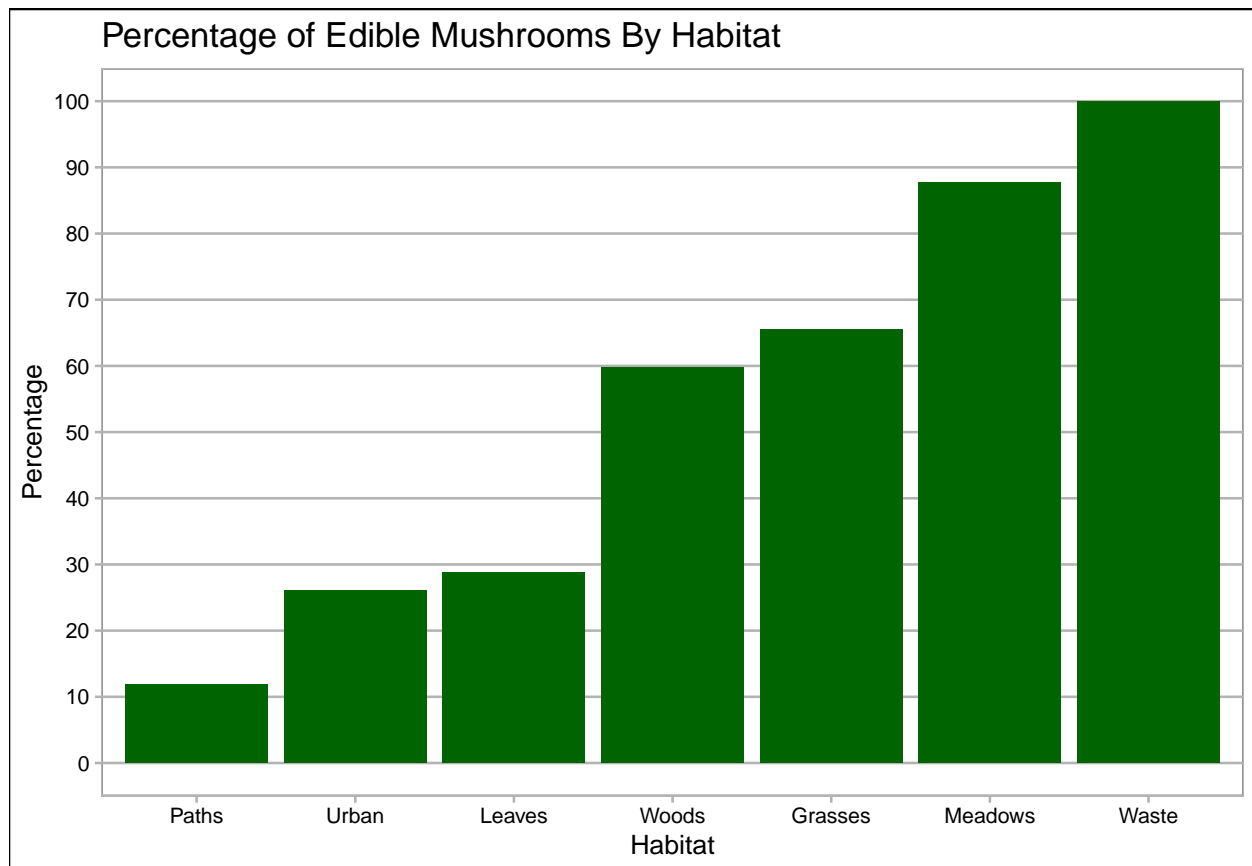
Models

In this section, we used the features of the mushrooms to predict the mushroom's edibility. We used logistic regression, k-nearest neighbors, and a classification tree to make the predictions.

Training & Test Sets

First, we split the data into a training set, which consisted of 80% of the data, and a test set, which consisted of the remaining 20%.

```
# Split the data set into a training set (80%) and a test set (20%)
set.seed(2)
test_index <- createDataPartition(data$edibility, times = 1, p = 0.2, list = FALSE)
train_set <- data[-test_index,]
test_set <- data[test_index,]
rm(test_index)
```

To ensure that the prevalence of poisonous and edible mushrooms are approximately the same, we checked each set. We found that the prevalences are approximately the same.

```
# Check the prevalence of edible mushrooms in each set
mean(data$edibility == "Edible")
```

```
## [1] 0.5179714
```

```
mean(train_set$edibility == "Edible")
```

```
## [1] 0.5180055
```

```
mean(test_set$edibility == "Edible")
```

```
## [1] 0.5178352
```

For the following models, the positive class was `Edible`. By predicting every mushroom to be edible, we would achieve an accuracy of 0.5178352. This would be the baseline model. However, the goal was to improve the accuracy.

Logistic Regression

The first model used was the logistic regression model. Surprisingly, we managed to achieve a 100% accuracy with this model.

```
# Train the model
set.seed(2)
train_glm <- train(edibility ~ ., method = "glm", data = train_set)

# Predict the outcomes
y_hat_glm <- predict(train_glm, test_set)

# Compute the results
confusionMatrix(y_hat_glm, test_set$edibility)
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction  Edible Poisonous
##    Edible      842         0
##    Poisonous     0       784
##
##               Accuracy : 1
##                 95% CI : (0.9977, 1)
##     No Information Rate : 0.5178
##     P-Value [Acc > NIR] : < 0.00000000000000022
##
##                  Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
##            Sensitivity : 1.0000
##            Specificity : 1.0000
##         Pos Pred Value : 1.0000
##         Neg Pred Value : 1.0000
##             Prevalence : 0.5178
##         Detection Rate : 0.5178
```

```
##    Detection Prevalence : 0.5178
##        Balanced Accuracy : 1.0000
##
##           'Positive' Class : Edible
##
```

K-Nearest Neighbors

The next model used was the k-nearest neighbors model. It also managed to achieve a 100% accuracy. However, this algorithm took the longest to complete.

```r
# Train the model
set.seed(2)
train_knn <- train(edibility ~ ., method = "knn", data = train_set,
                   tuneGrid = data.frame(k = 2:7))

# Predict the outcomes
y_hat_knn <- predict(train_knn, test_set)

# Compute the results
confusionMatrix(y_hat_knn, test_set$edibility)
```
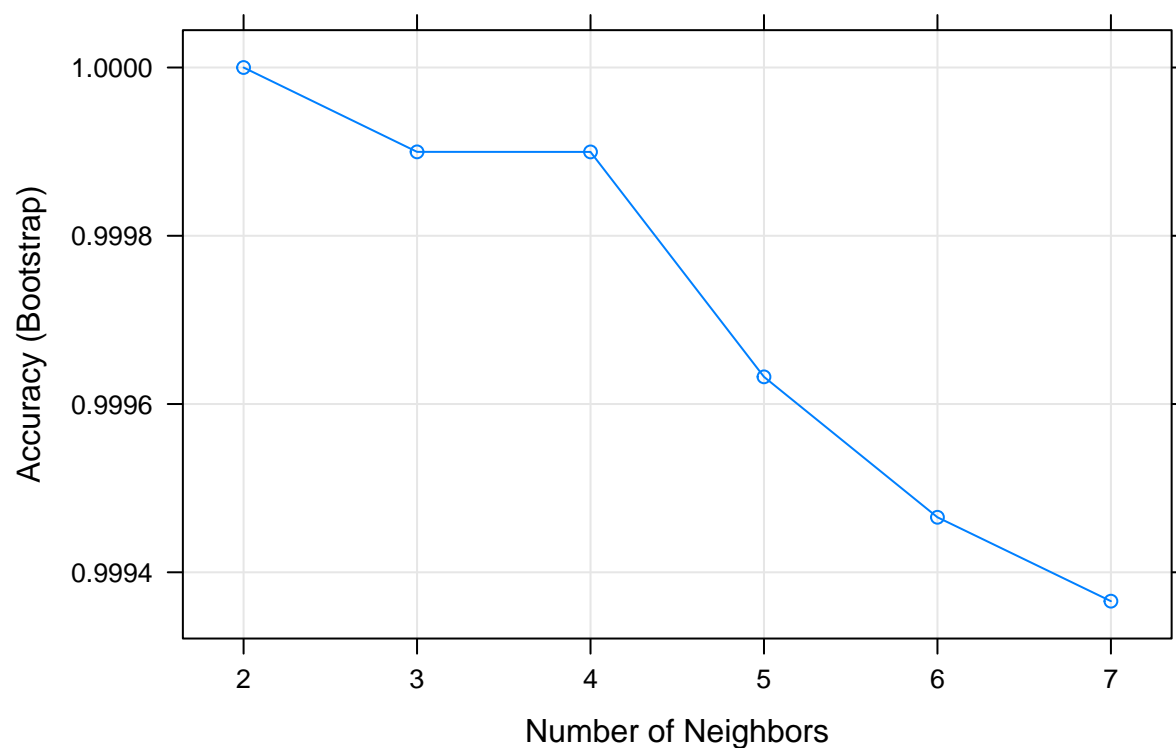
```
## Confusion Matrix and Statistics
##
##              Reference
## Prediction  Edible Poisonous
##    Edible       842         0
##    Poisonous      0       784
##
##                  Accuracy : 1
##                    95% CI : (0.9977, 1)
##       No Information Rate : 0.5178
##       P-Value [Acc > NIR] : < 0.00000000000000022
##
##                     Kappa : 1
##
##   Mcnemar's Test P-Value : NA
##
##               Sensitivity : 1.0000
##               Specificity : 1.0000
##            Pos Pred Value : 1.0000
##            Neg Pred Value : 1.0000
##                Prevalence : 0.5178
##            Detection Rate : 0.5178
##      Detection Prevalence : 0.5178
##         Balanced Accuracy : 1.0000
##
##           'Positive' Class : Edible
##
```

We can see the accuracies among different values of k, where the most optimal value was 2.

K–Nearest Neighbors Results

Classification Tree

The final model used was the classification tree. This model was able to achieve a 100% accuracy as well.

```
# Train the model
set.seed(2)
train_ct <- train(edibility ~ ., method = "rpart", data = train_set,
                  tuneGrid = data.frame(cp = seq(0, 0.001, 0.0001)))

# Predict the outcomes
y_hat_ct <- predict(train_ct, test_set)

# Compute the results
confusionMatrix(y_hat_ct, test_set$edibility)
```

```
## Confusion Matrix and Statistics
##
##            Reference
## Prediction  Edible Poisonous
##   Edible        842        0
##   Poisonous       0      784
##
##                Accuracy : 1
##                  95% CI : (0.9977, 1)
##     No Information Rate : 0.5178
```

20

```
##      P-Value [Acc > NIR] : < 0.0000000000000022
##
##                    Kappa : 1
##
##   Mcnemar's Test P-Value : NA
##
##              Sensitivity : 1.0000
##              Specificity : 1.0000
##           Pos Pred Value : 1.0000
##           Neg Pred Value : 1.0000
##               Prevalence : 0.5178
##           Detection Rate : 0.5178
##     Detection Prevalence : 0.5178
##         Balanced Accuracy : 1.0000
##
##         'Positive' Class : Edible
##
```

The following graph shows the accuracies using different complexity parameters. According to the model, the most optimal paramater was 0.0003.

Classification Tree Results
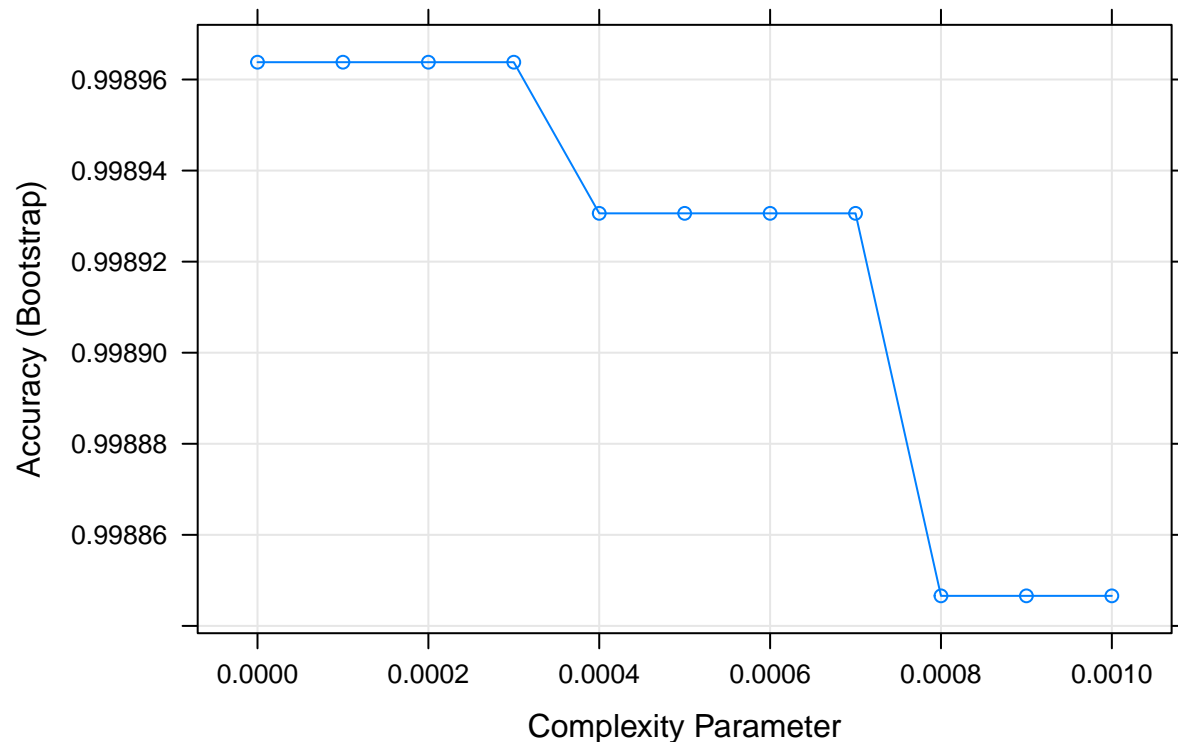
We can also observe the most important variables in the model. It appears that odor and stalk surface were among one of the most important variables. Having no odor was the most important variable.

```
# Show the most important variables according to the model
varImp(train_ct)
```

```
## rpart variable importance
##
##   only 20 most important variables shown (out of 95)
##
##                                 Overall
## odorNone                        100.000
## odorFoul                         62.643
## stalk.surface.above.ringSilky    59.503
## stalk.surface.below.ringSilky    53.450
## gill.sizeNarrow                  49.316
## bruisesTRUE                      40.436
## odorAnise                        38.499
## ring.typePendant                 29.422
## stalk.rootClub                   28.566
## stalk.surface.below.ringScaly    21.372
## habitatMeadows                   17.920
## stalk.rootRooted                 14.414
## spore.print.colorGreen            5.544
## gill.spacingCrowded               5.396
## spore.print.colorPurple           5.341
## cap.colorYellow                   3.464
## odorMusty                         2.977
## stalk.surface.above.ringSmooth    2.977
## stalk.color.below.ringYellow      1.768
## gill.colorGreen                   1.696
```

Results

Here are the results from the three models above:

```
##                  Model Accuracy
## 1 Logistic Regression      100
## 2 K-Nearest Neighbors      100
## 3 Classification Tree      100
```

Interestingly, all 3 models achieved a **100% accuracy**.

Conclusion

While there are several features that can be used to identify edible or poisonous mushrooms, odor seems to be the best indicator of edibility. Those with a scent can be identified as edible or poisonous while those with no odor were mostly safe to eat. Stalk color also seem to be a good indicator as well.

Regarding the models, all three models perfectly predicted mushroom edibility. However, a limitation of the data set is that some of the column values were not as prevalent. For example, there were 4 mushrooms that had a conical cap shape, 16 with a purple cap color, 8 with a yellow stalk color above their ring, etc. Perhaps more data would've been beneficial as there may have been anomalies not captured by the data set.