

Joe_Rovalino_Data607_wk#1

Joe Rovalino

8/29/2019

R Markdown

Relevant Information: This data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family (pp. 500-525). Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. The Guide clearly states that there is no simple rule for determining the edibility of a mushroom; no rule like “leaflets three, let it be” for Poisonous Oak and Ivy.

```
df <- read.table("https://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/agaricus-lepiota.d
                header = FALSE, sep=",", stringsAsFactors=FALSE)
#nrow(df)
#ncol(df)
str(df)
```

```
## 'data.frame': 8124 obs. of 23 variables:
```

```
## $ V1 : chr "p" "e" "e" "p" ...
## $ V2 : chr "x" "x" "b" "x" ...
## $ V3 : chr "s" "s" "s" "y" ...
## $ V4 : chr "n" "y" "w" "w" ...
## $ V5 : chr "t" "t" "t" "t" ...
## $ V6 : chr "p" "a" "l" "p" ...
## $ V7 : chr "f" "f" "f" "f" ...
## $ V8 : chr "c" "c" "c" "c" ...
## $ V9 : chr "n" "b" "b" "n" ...
## $ V10: chr "k" "k" "n" "n" ...
## $ V11: chr "e" "e" "e" "e" ...
## $ V12: chr "e" "c" "c" "e" ...
## $ V13: chr "s" "s" "s" "s" ...
## $ V14: chr "s" "s" "s" "s" ...
## $ V15: chr "w" "w" "w" "w" ...
## $ V16: chr "w" "w" "w" "w" ...
## $ V17: chr "p" "p" "p" "p" ...
## $ V18: chr "w" "w" "w" "w" ...
## $ V19: chr "o" "o" "o" "o" ...
## $ V20: chr "p" "p" "p" "p" ...
## $ V21: chr "k" "n" "n" "k" ...
## $ V22: chr "s" "n" "n" "s" ...
## $ V23: chr "u" "g" "m" "u" ...
```

```
#head(df)
#summary(df)
dim(df)
```

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

```
colnames(df)
```

```
## [1] "V1" "V2" "V3" "V4" "V5" "V6" "V7" "V8" "V9" "V10" "V11"  
## [12] "V12" "V13" "V14" "V15" "V16" "V17" "V18" "V19" "V20" "V21" "V22"  
## [23] "V23"
```

Going to create a subset DF with better names. First going to change the columns names to something more descriptive than V#. Found the example to do this on following: <https://www.datanovia.com/en/lessons/rename-data-frame-columns-in-r/>

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1      v purrr  0.3.2  
## v tibble  2.1.3      v dplyr  0.8.3  
## v tidyr   0.8.3      v stringr 1.4.0  
## v readr   1.3.1      v forcats 0.4.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()
```

```
subdf <- df %>%  
  rename(  
    EDIBLE = V1,  
    CONVEX = V2,  
    SMOOTH = V3,  
    WHITE = V4,  
    BRUISES = V5,  
    ALMOND = V6,  
    FREE = V7,  
    CROWDED = V8,  
    NARROW = V9,  
    WHITE = V10,  
    TAPERING = V11,  
    BULBOUS = V12,  
    SMOOTH = V13,  
    SMOOTH = V14,  
    WHITE = V15,  
    WHITE = V16,  
    PARTIAL = V17,  
    WHITE = V18,  
    ONE = V19,  
    PENDANT = V20,  
    PURPLE = V21,  
    SEVERAL = V22,  
    WOODS = V23  
  )
```

Next step: You should include the column that indicates edible or poisonous and three or four other columns. The rename step was already performed above using `dplyr::rename()`. So now will create the dataframe with 4 columns inclusive of Edible

```
Four_col_df <- subdf[, c(1,3,5,7)]  
head(Four_col_df)
```

```
##    EDIBLE SMOOTH BRUISES FREE  
## 1      p      s      t      f  
## 2      e      s      t      f  
## 3      e      s      t      f  
## 4      p      y      t      f  
## 5      e      s      f      f  
## 6      e      y      t      f
```