HW2

Summary of Mushroom

Ting-Ni, Chen

2025-03-20

目錄

1. Variable Definition

Variable	Data Type	Definition
family	character	The family of the mushroom.
name	character	The name of the mushroom.
class	categorical	edible=e, poisonous=p.
cap-diameter (m)	numerical	Number(s) in cm. Two values = min max, one value = mean.
cap-shape (n)	categorical	bell = b, conical = c, convex = x, flat = f, sunken = s, spherical = p, others = o.
cap-surface (n)	categorical	fibrous = i, grooves = g, scaly = y, smooth = s, shiny = h, leathery = l, silky = k, sticky = t, wrinkled = w, fleshy = e.
cap-color (n)	categorical	brown = n, buff = b, gray = g, green = r, pink = p, purple = u, red = e, white = w, yellow = y, blue = l, orange = o, black = k.
does-bruise-bleed (n)	categorical	bruises-or-bleeding = t, no = f.
gill-attachment (n)	categorical	adnate = a, adnexed = x, decurrent = d, free = e, sinuate = s, pores = p, none = f, unknown = ?.
gill-spacing (n)	categorical	close = c, distant = d, none = f.
gill-color (n)	categorical	see cap-color + none = f.
stem-height (m)	numerical	Number(s) in cm. Two values = min max, one value = mean.
stem-width (m)	numerical	Number(s) in mm. Two values = min max, one value = mean.
stem-root (n)	categorical	bulbous = b, swollen = s, club = c, cup = u, equal = e, rhizomorphs = z, rooted = r.
stem-surface (n)	categorical	see cap-surface + none = f.
stem-color (n)	categorical	see cap-color + none = f.
veil-type (n)	categorical	partial = p, universal = u.

veil-color (n)	categorical	see cap-color + none = f.
has-ring (n)	categorical	ring = t, none = f.
ring-type (n)	categorical	cobwebby = c, evanescent = e, flaring = r, grooved = g, large = l, pendant = p, sheathing = s, zone = z, scaly = y, movable = m, none = f, unknown = ?.
spore-print-color (n)	categorical	see cap-color.
habitat (n)	categorical	grasses = g, leaves = l, meadows = m, paths = p, heaths = h, urban = u, waste = w, woods = d.
season (n)	categorical	spring = s, summer = u, autumn = a, winter = w.

表 1: Mushroom Data Dictionary

2. One Hot Encoding

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MultiLabelBinarizer

df = pd.read_csv('./mushroom/primary_data.csv', sep=';')

df.replace("NA", np.nan, inplace=True)

df_encoded = df.copy()

categorical_columns = ["cap-shape", "Cap-surface", "cap-color", "does-bruise-or-bleed",
                       "gill-attachment", "gill-spacing", "gill-color", "stem-root",
                       "stem-surface", "stem-color", "veil-type", "veil-color", "has-ring",
                       "ring-type", "Spore-print-color", "habitat", "season"]

def one_hot_encode_multilabel(df, columns):
    df_encoded = df.copy()
    mlb = MultiLabelBinarizer()

    for col in columns:
        if col in df_encoded.columns:

            df_encoded[col] = df_encoded[col].fillna("MISSING").astype(str)
            temp_col = df_encoded[col].apply(lambda x: x.strip("[]") if isinstance(x, str) else x)
            df_encoded[col] = temp_col.apply(lambda x: x.split(", ") if x != "MISSING" else [])
            one_hot = pd.DataFrame(mlb.fit_transform(df_encoded[col]),
                                   columns=[f"{col}_{cls}" for cls in mlb.classes_])
            one_hot[df_encoded[col].apply(lambda x: x == [])] = pd.NA
            df_encoded = df_encoded.drop(columns=[col]).join(one_hot)

    return df_encoded

df_one_hot = one_hot_encode_multilabel(df_encoded, categorical_columns)
```

3. Data discription

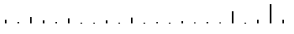
```
library(Hmisc)
library(dplyr)
library(reticulate)
library(stringr)

df <- py_to_r(py$df_one_hot)
df[df == "NaN"] <- NA
df <- data.frame(lapply(df, as.factor))
desc_stats <- describe(df)

latex(desc_stats, descript = "Descriptive Statistics", file = '', caption.placement = "top")
```

df
124 Variables 173 Observations

family



n	missing	distinct
173	0	23

```
lowest : Amanita Family     Bolbitius Family   Bolete Family      Bracket Fungi      Chanterelle Family
highest: Russula Family     Saddle-Cup Family  Stropharia Family  Tricholoma Family  Wax Gill Family
```

name

n	missing	distinct
173	0	173

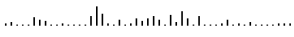
```
lowest : Amethyst Deceiver        Aniseed Funnel Cap       Apricot Fungus           Bare-toothed Russula     Bay Bolete
highest: Yellow-gilled Russula    Yellow-staining Mushroom Yellow-stemmed Bell Cap  Yellow Swamp Russula     Yellow Wax cap
```

class

n	missing	distinct
173	0	2

Value	e	p
Frequency	77	96
Proportion	0.445	0.555

cap.diameter



n	missing	distinct
173	0	51

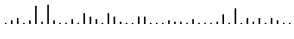
```
lowest : [0.4, 1]   [0.5, 1.5] [0.5, 1]   [0.7, 1.3] [1, 1.5]
highest: [8, 14]    [8, 15]    [8, 20]    [8, 25]    [8, 30]
```

stem.height

n	missing	distinct
173	0	46

```
lowest : [0]      [1, 2]   [1, 3]   [10, 12] [10, 15], highest: [8, 12]  [8, 15]  [8, 20]  [8, 25]  [8, 30]
```

stem.width



n	missing	distinct
173	0	48

```
lowest : [0.5, 1] [0]      [1, 2]   [1, 3]   [1]     , highest: [7, 15]  [8, 12]  [8, 15]  [8, 18]  [8, 20]
```

cap.shape_b

n	missing	distinct
173	0	2

Value	0	1
Frequency	150	23
Proportion	0.867	0.133

cap.shape_c

n	missing	distinct
173	0	2

Value	0	1
Frequency	165	8
Proportion	0.954	0.046

---

cap.shape_f

n	missing	distinct
173	0	2

Value	0	1
Frequency	99	74
Proportion	0.572	0.428

---

cap.shape_o

n	missing	distinct
173	0	2

Value	0	1
Frequency	161	12
Proportion	0.931	0.069

---

cap.shape_p

n	missing	distinct
173	0	2

Value	0	1
Frequency	158	15
Proportion	0.913	0.087

---

cap.shape_s

n	missing	distinct
173	0	2

Value	0	1
Frequency	137	36
Proportion	0.792	0.208

---

cap.shape_x

n	missing	distinct
173	0	2

Value	0	1
Frequency	63	110
Proportion	0.364	0.636

Cap.surface_d

n	missing	distinct
133	40	2

Value	0	1
Frequency	115	18
Proportion	0.865	0.135

Cap.surface_e

n	missing	distinct
133	40	2

Value	0	1
Frequency	122	11
Proportion	0.917	0.083

Cap.surface_g

n	missing	distinct
133	40	2

Value	0	1
Frequency	117	16
Proportion	0.88	0.12

Cap.surface_h

n	missing	distinct
133	40	2

Value	0	1
Frequency	107	26
Proportion	0.805	0.195

Cap.surface_i

n	missing	distinct
133	40	2

Value	0	1
Frequency	124	9
Proportion	0.932	0.068

Cap.surface_k

n	missing	distinct
133	40	2

Value	0	1
Frequency	123	10
Proportion	0.925	0.075

Cap.surface_l

n	missing	distinct
133	40	2

Value	0	1
Frequency	129	4
Proportion	0.97	0.03

Cap.surface_s

n	missing	distinct
133	40	2

Value	0	1
Frequency	100	33
Proportion	0.752	0.248

Cap.surface_t

n	missing	distinct
133	40	2

Value	0	1
Frequency	96	37
Proportion	0.722	0.278

Cap.surface_w

n	missing	distinct
133	40	2

Value	0	1
Frequency	125	8
Proportion	0.94	0.06

Cap.surface_y

n	missing	distinct
133	40	2

Value	0	1
Frequency	110	23
Proportion	0.827	0.173

---

cap.color_b

n	missing	distinct
173	0	2

Value	0	1
Frequency	166	7
Proportion	0.96	0.04

---

cap.color_e

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    148    25
Proportion 0.855 0.145
```

---

cap.color_g

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    145    28
Proportion 0.838 0.162
```

---

cap.color_k

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    164     9
Proportion 0.948 0.052
```

---

cap.color_l

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    167     6
Proportion 0.965 0.035
```

---

cap.color_n

```
   n    missing    distinct
 173          0          2

Value         0    1
Frequency    64  109
Proportion 0.37 0.63
```

---

cap.color_n..w

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    172     1
Proportion 0.994 0.006
```

---

cap.color_o

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    151    22
Proportion 0.873 0.127
```

---

cap.color_p

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    162    11
Proportion 0.936 0.064
```

---

cap.color_r

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    160    13
Proportion 0.925 0.075
```

---

cap.color_u

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    163    10
Proportion 0.942 0.058
```

---

cap.color_w

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    139    34
Proportion 0.803 0.197
```

---

cap.color_y

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    129    44
Proportion 0.746 0.254
```

---

does.bruise.or.bleed_f

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency     30   143
Proportion 0.173 0.827
```

---

does.bruise.or.bleed_t

```
   n    missing    distinct
 173          0          2

Value          0     1
Frequency    143    30
Proportion 0.827 0.173
```

---

gill.attachment_a

```
   n    missing    distinct
 145         28          2

Value          0     1
Frequency    105    40
Proportion 0.724 0.276
```

---

gill.attachment_d

```
   n    missing    distinct
 145         28          2

Value          0     1
Frequency    112    33
Proportion 0.772 0.228
```

---

gill.attachment_e

```
   n    missing    distinct
 145         28          2

Value         0    1
Frequency   129   16
Proportion 0.89 0.11
```

---

gill.attachment_f

```
   n    missing    distinct
 145         28          2

Value          0     1
Frequency    135    10
Proportion 0.931 0.069
```

---

gill.attachment_p

n	missing	distinct
145	28	2

Value	0	1
Frequency	128	17
Proportion	0.883	0.117

---

gill.attachment_s

n	missing	distinct
145	28	2

Value	0	1
Frequency	129	16
Proportion	0.89	0.11

---

gill.attachment_x

n	missing	distinct
145	28	2

Value	0	1
Frequency	124	21
Proportion	0.855	0.145

---

gill.spacing_c

n	missing	distinct
102	71	2

Value	0	1
Frequency	32	70
Proportion	0.314	0.686

---

gill.spacing_d

n	missing	distinct
102	71	2

Value	0	1
Frequency	80	22
Proportion	0.784	0.216

---

gill.spacing_f

n	missing	distinct
102	71	2

Value	0	1
Frequency	92	10
Proportion	0.902	0.098

---

gill.color_b

n	missing	distinct
173	0	2

Value	0	1
Frequency	168	5
Proportion	0.971	0.029

---

gill.color_e

n	missing	distinct
173	0	2

Value	0	1
Frequency	167	6
Proportion	0.965	0.035

---

gill.color_f

n	missing	distinct
173	0	2

Value	0	1
Frequency	163	10
Proportion	0.942	0.058

---

gill.color_g

```
   n   missing   distinct
 173         0         2

Value          0     1
Frequency    150    23
Proportion 0.867 0.133
```

---

gill.color_k

```
   n   missing   distinct
 173         0         2

Value          0     1
Frequency    158    15
Proportion 0.913 0.087
```

---

gill.color_n

```
   n   missing   distinct
 173         0         2

Value          0     1
Frequency    126    47
Proportion 0.728 0.272
```

---

gill.color_o

```
   n   missing   distinct
 173         0         2

Value          0     1
Frequency    160    13
Proportion 0.925 0.075
```

---

gill.color_p

```
   n   missing   distinct
 173         0         2

Value          0     1
Frequency    145    28
Proportion 0.838 0.162
```

---

gill.color_r

```
   n   missing   distinct
 173         0         2

Value          0     1
Frequency    165     8
Proportion 0.954 0.046
```

---

gill.color_u

```
   n   missing   distinct
 173         0         2

Value         0    1
Frequency   166    7
Proportion 0.96 0.04
```

---

gill.color_w

```
   n   missing   distinct
 173         0         2

Value          0     1
Frequency    100    73
Proportion 0.578 0.422
```

---

gill.color_y

```
   n   missing   distinct
 173         0         2

Value          0     1
Frequency    129    44
Proportion 0.746 0.254
```

---

stem.root_b

n	missing	distinct
27	146	2

Value	0	1
Frequency	18	9
Proportion	0.667	0.333

---

stem.root_c

n	missing	distinct
27	146	2

Value	0	1
Frequency	25	2
Proportion	0.926	0.074

---

stem.root_f

n	missing	distinct
27	146	2

Value	0	1
Frequency	24	3
Proportion	0.889	0.111

---

stem.root_r

n	missing	distinct
27	146	2

Value	0	1
Frequency	23	4
Proportion	0.852	0.148

---

stem.root_s

n	missing	distinct
27	146	2

Value	0	1
Frequency	18	9
Proportion	0.667	0.333

---

stem.surface_f

n	missing	distinct
65	108	2

Value	0	1
Frequency	62	3
Proportion	0.954	0.046

---

stem.surface_g

n	missing	distinct
65	108	2

Value	0	1
Frequency	60	5
Proportion	0.923	0.077

---

stem.surface_h

n	missing	distinct
65	108	2

Value	0	1
Frequency	63	2
Proportion	0.969	0.031

---

stem.surface_i

n	missing	distinct
65	108	2

Value	0	1
Frequency	51	14
Proportion	0.785	0.215

---

stem.surface_k

n	missing	distinct
65	108	2

Value	0	1
Frequency	60	5
Proportion	0.923	0.077

---

stem.surface_s

n	missing	distinct
65	108	2

Value	0	1
Frequency	46	19
Proportion	0.708	0.292

---

stem.surface_t

n	missing	distinct
65	108	2

Value	0	1
Frequency	57	8
Proportion	0.877	0.123

---

stem.surface_y

n	missing	distinct
65	108	2

Value	0	1
Frequency	50	15
Proportion	0.769	0.231

---

stem.color_.w

n	missing	distinct
173	0	2

Value	0	1
Frequency	172	1
Proportion	0.994	0.006

---

stem.color_b

n	missing	distinct
173	0	2

Value	0	1
Frequency	172	1
Proportion	0.994	0.006

---

stem.color_e

n	missing	distinct
173	0	2

Value	0	1
Frequency	162	11
Proportion	0.936	0.064

---

stem.color_f

n	missing	distinct
173	0	2

Value	0	1
Frequency	170	3
Proportion	0.983	0.017

---

stem.color_g

n	missing	distinct
173	0	2

Value	0	1
Frequency	159	14
Proportion	0.919	0.081

---

stem.color_k

n	missing	distinct
173	0	2

Value	0	1
Frequency	169	4
Proportion	0.977	0.023

---

stem.color_l

n	missing	distinct
173	0	2

Value	0	1
Frequency	171	2
Proportion	0.988	0.012

---

stem.color_n

n	missing	distinct
173	0	2

Value	0	1
Frequency	103	70
Proportion	0.595	0.405

---

stem.color_o

n	missing	distinct
173	0	2

Value	0	1
Frequency	161	12
Proportion	0.931	0.069

---

stem.color_p

n	missing	distinct
173	0	2

Value	0	1
Frequency	169	4
Proportion	0.977	0.023

---

stem.color_r

n	missing	distinct
173	0	2

Value	0	1
Frequency	169	4
Proportion	0.977	0.023

---

stem.color_u

n	missing	distinct
173	0	2

Value	0	1
Frequency	166	7
Proportion	0.96	0.04

---

stem.color_w

n	missing	distinct
173	0	2

Value	0	1
Frequency	100	73
Proportion	0.578	0.422

---

stem.color_y

n	missing	distinct
173	0	2

Value	0	1
Frequency	141	32
Proportion	0.815	0.185

---

veil.type_u

n	missing	distinct	value
9	164	1	1

Value	1
Frequency	9
Proportion	1

---

veil.color_e

n	missing	distinct
21	152	2

Value	0	1
Frequency	20	1
Proportion	0.952	0.048

---

veil.color_k

n	missing	distinct
21	152	2

Value	0	1
Frequency	20	1
Proportion	0.952	0.048

---

veil.color_n

n	missing	distinct
21	152	2

Value	0	1
Frequency	19	2
Proportion	0.905	0.095

---

veil.color_u

n	missing	distinct
21	152	2

Value	0	1
Frequency	20	1
Proportion	0.952	0.048

---

veil.color_w

n	missing	distinct
21	152	2

Value	0	1
Frequency	5	16
Proportion	0.238	0.762

---

veil.color_y

n	missing	distinct
21	152	2

Value	0	1
Frequency	19	2
Proportion	0.905	0.095

---

has.ring_f

n	missing	distinct
173	0	2

Value	0	1
Frequency	43	130
Proportion	0.249	0.751

---

has.ring_t

n	missing	distinct
173	0	2

Value	0	1
Frequency	130	43
Proportion	0.751	0.249

---

ring.type_e

```
   n    missing    distinct
 166          7          2

Value          0     1
Frequency    158     8
Proportion 0.952 0.048
```

---

ring.type_f

```
   n    missing    distinct
 166          7          2

Value          0     1
Frequency     29   137
Proportion 0.175 0.825
```

---

ring.type_g

```
   n    missing    distinct
 166          7          2

Value         0    1
Frequency   161    5
Proportion 0.97 0.03
```

---

ring.type_l

```
   n    missing    distinct
 166          7          2

Value          0     1
Frequency    160     6
Proportion 0.964 0.036
```

---

ring.type_m

```
   n    missing    distinct
 166          7          2

Value          0     1
Frequency    165     1
Proportion 0.994 0.006
```

---

ring.type_p

```
   n    missing    distinct
 166          7          2

Value         0    1
Frequency   161    5
Proportion 0.97 0.03
```

---

ring.type_r

```
   n    missing    distinct
 166          7          2

Value         0    1
Frequency   161    5
Proportion 0.97 0.03
```

---

ring.type_z

```
   n    missing    distinct
 166          7          2

Value          0     1
Frequency    160     6
Proportion 0.964 0.036
```

Spore.print.color_g

```
   n    missing    distinct
  18        155          2

Value          0     1
Frequency     17     1
Proportion 0.944 0.056
```

Spore.print.color_k

n	missing	distinct
18	155	2

Value	0	1
Frequency	11	7
Proportion	0.611	0.389

Spore.print.color_n

n	missing	distinct
18	155	2

Value	0	1
Frequency	15	3
Proportion	0.833	0.167

Spore.print.color_p

n	missing	distinct
18	155	2

Value	0	1
Frequency	14	4
Proportion	0.778	0.222

Spore.print.color_r

n	missing	distinct
18	155	2

Value	0	1
Frequency	17	1
Proportion	0.944	0.056

Spore.print.color_u

n	missing	distinct
18	155	2

Value	0	1
Frequency	17	1
Proportion	0.944	0.056

Spore.print.color_w

n	missing	distinct
18	155	2

Value	0	1
Frequency	14	4
Proportion	0.778	0.222

---

habitat_d

n	missing	distinct
173	0	2

Value	0	1
Frequency	22	151
Proportion	0.127	0.873

---

habitat_g

n	missing	distinct
173	0	2

Value	0	1
Frequency	135	38
Proportion	0.78	0.22

---

habitat_h

n	missing	distinct
173	0	2

Value	0	1
Frequency	160	13
Proportion	0.925	0.075

---

habitat_l

n	missing	distinct
173	0	2

Value	0	1
Frequency	155	18
Proportion	0.896	0.104

---

habitat_m

n	missing	distinct
173	0	2

Value	0	1
Frequency	156	17
Proportion	0.902	0.098

---

habitat_p

n	missing	distinct
173	0	2

Value	0	1
Frequency	171	2
Proportion	0.988	0.012

---

habitat_u

n	missing	distinct
173	0	2

Value	0	1
Frequency	172	1
Proportion	0.994	0.006

---

habitat_w

n	missing	distinct
173	0	2

Value	0	1
Frequency	172	1
Proportion	0.994	0.006

---

season_a

n	missing	distinct
173	0	2

Value	0	1
Frequency	5	168
Proportion	0.029	0.971

---

season_s

n	missing	distinct
173	0	2

Value	0	1
Frequency	150	23
Proportion	0.867	0.133

---

season_u

n	missing	distinct
173	0	2

Value	0	1
Frequency	33	140
Proportion	0.191	0.809

---

season_w

n	missing	distinct
173	0	2

Value	0	1
Frequency	132	41
Proportion	0.763	0.237

---

4. Table One

```
library(table1)
library(tidyr)
library(knitr)
library(kableExtra)

vars_to_split <- c("cap.diameter", "stem.height", "stem.width")

for (var in vars_to_split) {
  if (var %in% names(df)) {
    df <- df %>%
      mutate(!!var := gsub("\\[|\\]", "", .data[[var]])) %>%
      mutate(!!var := ifelse(grepl(",", .data[[var]]), .data[[var]], paste(.data[[var]], .data[[var]], s
      separate(var, into = c(paste0(var, ".min"), paste0(var, ".max")), sep = ", ", convert = TRUE) %>%
      mutate(!!paste0(var, ".mean") := ifelse(.data[[paste0(var, ".min")]] == .data[[paste0(var, ".max")
                                              .data[[paste0(var, ".min")]], NA)) %>%
      mutate(!!paste0(var, ".max") := ifelse(!is.na(.data[[paste0(var, ".mean")]]), NA, .data[[paste0(va
  }
}

df$class <- as.factor(df$class)

numerical_vars <- c("cap.diameter.min", "cap.diameter.max", "cap.diameter.mean",
                    "stem.height.min", "stem.height.max", "stem.height.mean",
                    "stem.width.min", "stem.width.max", "stem.width.mean")

categorical_vars <- setdiff(names(df), c(numerical_vars, "family", "name"))

df[categorical_vars] <- lapply(df[categorical_vars], as.factor)

df_subset <- df[, c(numerical_vars, categorical_vars), drop = FALSE]

df_subset <- as.data.frame(df_subset)


df_subset[categorical_vars] <- lapply(df_subset[categorical_vars], as.factor)

t1 <- table1(~ . | class, data = df_subset)

kable(t1, format = "latex", booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(latex_options = c("repeat_header"))
```

	e	p	Overall
	(N=77)	(N=96)	(N=173)
cap.diameter.min			
Mean (SD)	4.75 (5.74)	3.47 (2.27)	4.04 (4.22)
Median [Min, Max]	4.00 [0.500, 50.0]	3.00 [0.400, 10.0]	3.00 [0.400, 50.0]
cap.diameter.max			
Mean (SD)	10.3 (5.76)	8.29 (5.58)	9.20 (5.73)

(continued)

	e	p	Overall
Median [Min, Max]	10.0 [1.50, 30.0]	7.00 [1.00, 30.0]	8.00 [1.00, 30.0]
Missing	1 (1.3%)	0 (0%)	1 (0.6%)
cap.diameter.mean			
Mean (SD)	50.0 (NA)	NA (NA)	50.0 (NA)
Median [Min, Max]	50.0 [50.0, 50.0]	NA [NA, NA]	50.0 [50.0, 50.0]
Missing	76 (98.7%)	96 (100%)	172 (99.4%)
stem.height.min			
Mean (SD)	4.52 (2.20)	4.14 (2.31)	4.31 (2.26)
Median [Min, Max]	4.00 [2.00, 15.0]	4.00 [0, 15.0]	4.00 [0, 15.0]
stem.height.max			
Mean (SD)	9.58 (5.03)	8.57 (3.80)	9.03 (4.41)
Median [Min, Max]	8.00 [3.00, 35.0]	8.00 [2.00, 20.0]	8.00 [2.00, 35.0]
Missing	0 (0%)	3 (3.1%)	3 (1.7%)
stem.height.mean			
Mean (SD)	NA (NA)	0 (0)	0 (0)
Median [Min, Max]	NA [NA, NA]	0 [0, 0]	0 [0, 0]
Missing	77 (100%)	93 (96.9%)	170 (98.3%)
stem.width.min			
Mean (SD)	10.1 (6.80)	7.26 (5.71)	8.53 (6.36)
Median [Min, Max]	10.0 [1.00, 40.0]	5.00 [0, 20.0]	8.00 [0, 40.0]
stem.width.max			
Mean (SD)	19.2 (15.9)	14.4 (11.8)	16.6 (13.9)
Median [Min, Max]	15.0 [2.00, 100]	10.0 [1.00, 60.0]	15.0 [1.00, 100]
Missing	4 (5.2%)	7 (7.3%)	11 (6.4%)
stem.width.mean			
Mean (SD)	7.75 (4.50)	2.00 (3.61)	4.09 (4.72)
Median [Min, Max]	10.0 [1.00, 10.0]	1.00 [0, 10.0]	1.00 [0, 10.0]
Missing	73 (94.8%)	89 (92.7%)	162 (93.6%)
cap.shape_b			
0	72 (93.5%)	78 (81.3%)	150 (86.7%)
1	5 (6.5%)	18 (18.8%)	23 (13.3%)
cap.shape_c			
0	73 (94.8%)	92 (95.8%)	165 (95.4%)
1	4 (5.2%)	4 (4.2%)	8 (4.6%)
cap.shape_f			
0	41 (53.2%)	58 (60.4%)	99 (57.2%)
1	36 (46.8%)	38 (39.6%)	74 (42.8%)

(continued)

	e	p	Overall
cap.shape_o			
0	73 (94.8%)	88 (91.7%)	161 (93.1%)
1	4 (5.2%)	8 (8.3%)	12 (6.9%)
cap.shape_p			
0	67 (87.0%)	91 (94.8%)	158 (91.3%)
1	10 (13.0%)	5 (5.2%)	15 (8.7%)
cap.shape_s			
0	60 (77.9%)	77 (80.2%)	137 (79.2%)
1	17 (22.1%)	19 (19.8%)	36 (20.8%)
cap.shape_x			
0	23 (29.9%)	40 (41.7%)	63 (36.4%)
1	54 (70.1%)	56 (58.3%)	110 (63.6%)
Cap.surface_d			
0	50 (64.9%)	65 (67.7%)	115 (66.5%)
1	8 (10.4%)	10 (10.4%)	18 (10.4%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_e			
0	54 (70.1%)	68 (70.8%)	122 (70.5%)
1	4 (5.2%)	7 (7.3%)	11 (6.4%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_g			
0	51 (66.2%)	66 (68.8%)	117 (67.6%)
1	7 (9.1%)	9 (9.4%)	16 (9.2%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_h			
0	45 (58.4%)	62 (64.6%)	107 (61.8%)
1	13 (16.9%)	13 (13.5%)	26 (15.0%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_i			
0	56 (72.7%)	68 (70.8%)	124 (71.7%)
1	2 (2.6%)	7 (7.3%)	9 (5.2%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_k			
0	57 (74.0%)	66 (68.8%)	123 (71.1%)
1	1 (1.3%)	9 (9.4%)	10 (5.8%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_l			

(continued)

	e	p	Overall
0	56 (72.7%)	73 (76.0%)	129 (74.6%)
1	2 (2.6%)	2 (2.1%)	4 (2.3%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_s			
0	40 (51.9%)	60 (62.5%)	100 (57.8%)
1	18 (23.4%)	15 (15.6%)	33 (19.1%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_t			
0	43 (55.8%)	53 (55.2%)	96 (55.5%)
1	15 (19.5%)	22 (22.9%)	37 (21.4%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_w			
0	55 (71.4%)	70 (72.9%)	125 (72.3%)
1	3 (3.9%)	5 (5.2%)	8 (4.6%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
Cap.surface_y			
0	46 (59.7%)	64 (66.7%)	110 (63.6%)
1	12 (15.6%)	11 (11.5%)	23 (13.3%)
Missing	19 (24.7%)	21 (21.9%)	40 (23.1%)
cap.color_b			
0	72 (93.5%)	94 (97.9%)	166 (96.0%)
1	5 (6.5%)	2 (2.1%)	7 (4.0%)
cap.color_e			
0	70 (90.9%)	78 (81.3%)	148 (85.5%)
1	7 (9.1%)	18 (18.8%)	25 (14.5%)
cap.color_g			
0	63 (81.8%)	82 (85.4%)	145 (83.8%)
1	14 (18.2%)	14 (14.6%)	28 (16.2%)
cap.color_k			
0	74 (96.1%)	90 (93.8%)	164 (94.8%)
1	3 (3.9%)	6 (6.3%)	9 (5.2%)
cap.color_l			
0	73 (94.8%)	94 (97.9%)	167 (96.5%)
1	4 (5.2%)	2 (2.1%)	6 (3.5%)
cap.color_n			
0	25 (32.5%)	39 (40.6%)	64 (37.0%)

(continued)

	e	p	Overall
1	52 (67.5%)	57 (59.4%)	109 (63.0%)
cap.color_n..w			
0	76 (98.7%)	96 (100%)	172 (99.4%)
1	1 (1.3%)	0 (0%)	1 (0.6%)
cap.color_o			
0	70 (90.9%)	81 (84.4%)	151 (87.3%)
1	7 (9.1%)	15 (15.6%)	22 (12.7%)
cap.color_p			
0	73 (94.8%)	89 (92.7%)	162 (93.6%)
1	4 (5.2%)	7 (7.3%)	11 (6.4%)
cap.color_r			
0	75 (97.4%)	85 (88.5%)	160 (92.5%)
1	2 (2.6%)	11 (11.5%)	13 (7.5%)
cap.color_u			
0	72 (93.5%)	91 (94.8%)	163 (94.2%)
1	5 (6.5%)	5 (5.2%)	10 (5.8%)
cap.color_w			
0	61 (79.2%)	78 (81.3%)	139 (80.3%)
1	16 (20.8%)	18 (18.8%)	34 (19.7%)
cap.color_y			
0	61 (79.2%)	68 (70.8%)	129 (74.6%)
1	16 (20.8%)	28 (29.2%)	44 (25.4%)
does.bruise.or.bleed_f			
0	14 (18.2%)	16 (16.7%)	30 (17.3%)
1	63 (81.8%)	80 (83.3%)	143 (82.7%)
does.bruise.or.bleed_t			
0	63 (81.8%)	80 (83.3%)	143 (82.7%)
1	14 (18.2%)	16 (16.7%)	30 (17.3%)
gill.attachment_a			
0	51 (66.2%)	54 (56.3%)	105 (60.7%)
1	16 (20.8%)	24 (25.0%)	40 (23.1%)
Missing	10 (13.0%)	18 (18.8%)	28 (16.2%)
gill.attachment_d			
0	53 (68.8%)	59 (61.5%)	112 (64.7%)
1	14 (18.2%)	19 (19.8%)	33 (19.1%)
Missing	10 (13.0%)	18 (18.8%)	28 (16.2%)
gill.attachment_e			

(continued)

	e	p	Overall
0	57 (74.0%)	72 (75.0%)	129 (74.6%)
1	10 (13.0%)	6 (6.3%)	16 (9.2%)
Missing	10 (13.0%)	18 (18.8%)	28 (16.2%)
gill.attachment_f			
0	63 (81.8%)	72 (75.0%)	135 (78.0%)
1	4 (5.2%)	6 (6.3%)	10 (5.8%)
Missing	10 (13.0%)	18 (18.8%)	28 (16.2%)
gill.attachment_p			
0	55 (71.4%)	73 (76.0%)	128 (74.0%)
1	12 (15.6%)	5 (5.2%)	17 (9.8%)
Missing	10 (13.0%)	18 (18.8%)	28 (16.2%)
gill.attachment_s			
0	60 (77.9%)	69 (71.9%)	129 (74.6%)
1	7 (9.1%)	9 (9.4%)	16 (9.2%)
Missing	10 (13.0%)	18 (18.8%)	28 (16.2%)
gill.attachment_x			
0	58 (75.3%)	66 (68.8%)	124 (71.7%)
1	9 (11.7%)	12 (12.5%)	21 (12.1%)
Missing	10 (13.0%)	18 (18.8%)	28 (16.2%)
gill.spacing_c			
0	17 (22.1%)	15 (15.6%)	32 (18.5%)
1	29 (37.7%)	41 (42.7%)	70 (40.5%)
Missing	31 (40.3%)	40 (41.7%)	71 (41.0%)
gill.spacing_d			
0	33 (42.9%)	47 (49.0%)	80 (46.2%)
1	13 (16.9%)	9 (9.4%)	22 (12.7%)
Missing	31 (40.3%)	40 (41.7%)	71 (41.0%)
gill.spacing_f			
0	42 (54.5%)	50 (52.1%)	92 (53.2%)
1	4 (5.2%)	6 (6.3%)	10 (5.8%)
Missing	31 (40.3%)	40 (41.7%)	71 (41.0%)
gill.color_b			
0	74 (96.1%)	94 (97.9%)	168 (97.1%)
1	3 (3.9%)	2 (2.1%)	5 (2.9%)
gill.color_e			
0	75 (97.4%)	92 (95.8%)	167 (96.5%)
1	2 (2.6%)	4 (4.2%)	6 (3.5%)

(continued)

	e	p	Overall
gill.color_f			
0	73 (94.8%)	90 (93.8%)	163 (94.2%)
1	4 (5.2%)	6 (6.3%)	10 (5.8%)
gill.color_g			
0	67 (87.0%)	83 (86.5%)	150 (86.7%)
1	10 (13.0%)	13 (13.5%)	23 (13.3%)
gill.color_k			
0	71 (92.2%)	87 (90.6%)	158 (91.3%)
1	6 (7.8%)	9 (9.4%)	15 (8.7%)
gill.color_n			
0	62 (80.5%)	64 (66.7%)	126 (72.8%)
1	15 (19.5%)	32 (33.3%)	47 (27.2%)
gill.color_o			
0	72 (93.5%)	88 (91.7%)	160 (92.5%)
1	5 (6.5%)	8 (8.3%)	13 (7.5%)
gill.color_p			
0	65 (84.4%)	80 (83.3%)	145 (83.8%)
1	12 (15.6%)	16 (16.7%)	28 (16.2%)
gill.color_r			
0	75 (97.4%)	90 (93.8%)	165 (95.4%)
1	2 (2.6%)	6 (6.3%)	8 (4.6%)
gill.color_u			
0	74 (96.1%)	92 (95.8%)	166 (96.0%)
1	3 (3.9%)	4 (4.2%)	7 (4.0%)
gill.color_w			
0	39 (50.6%)	61 (63.5%)	100 (57.8%)
1	38 (49.4%)	35 (36.5%)	73 (42.2%)
gill.color_y			
0	60 (77.9%)	69 (71.9%)	129 (74.6%)
1	17 (22.1%)	27 (28.1%)	44 (25.4%)
stem.root_b			
0	4 (5.2%)	14 (14.6%)	18 (10.4%)
1	6 (7.8%)	3 (3.1%)	9 (5.2%)
Missing	67 (87.0%)	79 (82.3%)	146 (84.4%)
stem.root_c			
0	10 (13.0%)	15 (15.6%)	25 (14.5%)

(continued)

	e	p	Overall
1	0 (0%)	2 (2.1%)	2 (1.2%)
Missing	67 (87.0%)	79 (82.3%)	146 (84.4%)
stem.root_f			
0	10 (13.0%)	14 (14.6%)	24 (13.9%)
1	0 (0%)	3 (3.1%)	3 (1.7%)
Missing	67 (87.0%)	79 (82.3%)	146 (84.4%)
stem.root_r			
0	10 (13.0%)	13 (13.5%)	23 (13.3%)
1	0 (0%)	4 (4.2%)	4 (2.3%)
Missing	67 (87.0%)	79 (82.3%)	146 (84.4%)
stem.root_s			
0	6 (7.8%)	12 (12.5%)	18 (10.4%)
1	4 (5.2%)	5 (5.2%)	9 (5.2%)
Missing	67 (87.0%)	79 (82.3%)	146 (84.4%)
stem.surface_f			
0	24 (31.2%)	38 (39.6%)	62 (35.8%)
1	0 (0%)	3 (3.1%)	3 (1.7%)
Missing	53 (68.8%)	55 (57.3%)	108 (62.4%)
stem.surface_g			
0	24 (31.2%)	36 (37.5%)	60 (34.7%)
1	0 (0%)	5 (5.2%)	5 (2.9%)
Missing	53 (68.8%)	55 (57.3%)	108 (62.4%)
stem.surface_h			
0	24 (31.2%)	39 (40.6%)	63 (36.4%)
1	0 (0%)	2 (2.1%)	2 (1.2%)
Missing	53 (68.8%)	55 (57.3%)	108 (62.4%)
stem.surface_i			
0	19 (24.7%)	32 (33.3%)	51 (29.5%)
1	5 (6.5%)	9 (9.4%)	14 (8.1%)
Missing	53 (68.8%)	55 (57.3%)	108 (62.4%)
stem.surface_k			
0	22 (28.6%)	38 (39.6%)	60 (34.7%)
1	2 (2.6%)	3 (3.1%)	5 (2.9%)
Missing	53 (68.8%)	55 (57.3%)	108 (62.4%)
stem.surface_s			
0	13 (16.9%)	33 (34.4%)	46 (26.6%)
1	11 (14.3%)	8 (8.3%)	19 (11.0%)

(continued)

	e	p	Overall
Missing	53 (68.8%)	55 (57.3%)	108 (62.4%)
stem.surface_t			
0	20 (26.0%)	37 (38.5%)	57 (32.9%)
1	4 (5.2%)	4 (4.2%)	8 (4.6%)
Missing	53 (68.8%)	55 (57.3%)	108 (62.4%)
stem.surface_y			
0	19 (24.7%)	31 (32.3%)	50 (28.9%)
1	5 (6.5%)	10 (10.4%)	15 (8.7%)
Missing	53 (68.8%)	55 (57.3%)	108 (62.4%)
stem.color_.w			
0	76 (98.7%)	96 (100%)	172 (99.4%)
1	1 (1.3%)	0 (0%)	1 (0.6%)
stem.color_b			
0	76 (98.7%)	96 (100%)	172 (99.4%)
1	1 (1.3%)	0 (0%)	1 (0.6%)
stem.color_e			
0	74 (96.1%)	88 (91.7%)	162 (93.6%)
1	3 (3.9%)	8 (8.3%)	11 (6.4%)
stem.color_f			
0	77 (100%)	93 (96.9%)	170 (98.3%)
1	0 (0%)	3 (3.1%)	3 (1.7%)
stem.color_g			
0	70 (90.9%)	89 (92.7%)	159 (91.9%)
1	7 (9.1%)	7 (7.3%)	14 (8.1%)
stem.color_k			
0	76 (98.7%)	93 (96.9%)	169 (97.7%)
1	1 (1.3%)	3 (3.1%)	4 (2.3%)
stem.color_l			
0	76 (98.7%)	95 (99.0%)	171 (98.8%)
1	1 (1.3%)	1 (1.0%)	2 (1.2%)
stem.color_n			
0	50 (64.9%)	53 (55.2%)	103 (59.5%)
1	27 (35.1%)	43 (44.8%)	70 (40.5%)
stem.color_o			
0	72 (93.5%)	89 (92.7%)	161 (93.1%)
1	5 (6.5%)	7 (7.3%)	12 (6.9%)
stem.color_p			

(continued)

	e	p	Overall
0	76 (98.7%)	93 (96.9%)	169 (97.7%)
1	1 (1.3%)	3 (3.1%)	4 (2.3%)
stem.color_r			
0	76 (98.7%)	93 (96.9%)	169 (97.7%)
1	1 (1.3%)	3 (3.1%)	4 (2.3%)
stem.color_u			
0	75 (97.4%)	91 (94.8%)	166 (96.0%)
1	2 (2.6%)	5 (5.2%)	7 (4.0%)
stem.color_w			
0	35 (45.5%)	65 (67.7%)	100 (57.8%)
1	42 (54.5%)	31 (32.3%)	73 (42.2%)
stem.color_y			
0	68 (88.3%)	73 (76.0%)	141 (81.5%)
1	9 (11.7%)	23 (24.0%)	32 (18.5%)
veil.type_u			
1	3 (3.9%)	6 (6.3%)	9 (5.2%)
Missing	74 (96.1%)	90 (93.8%)	164 (94.8%)
veil.color_e			
0	9 (11.7%)	11 (11.5%)	20 (11.6%)
1	0 (0%)	1 (1.0%)	1 (0.6%)
Missing	68 (88.3%)	84 (87.5%)	152 (87.9%)
veil.color_k			
0	9 (11.7%)	11 (11.5%)	20 (11.6%)
1	0 (0%)	1 (1.0%)	1 (0.6%)
Missing	68 (88.3%)	84 (87.5%)	152 (87.9%)
veil.color_n			
0	9 (11.7%)	10 (10.4%)	19 (11.0%)
1	0 (0%)	2 (2.1%)	2 (1.2%)
Missing	68 (88.3%)	84 (87.5%)	152 (87.9%)
veil.color_u			
0	9 (11.7%)	11 (11.5%)	20 (11.6%)
1	0 (0%)	1 (1.0%)	1 (0.6%)
Missing	68 (88.3%)	84 (87.5%)	152 (87.9%)
veil.color_w			
0	1 (1.3%)	4 (4.2%)	5 (2.9%)
1	8 (10.4%)	8 (8.3%)	16 (9.2%)

(continued)

	e	p	Overall
Missing	68 (88.3%)	84 (87.5%)	152 (87.9%)
veil.color_y			
0	7 (9.1%)	12 (12.5%)	19 (11.0%)
1	2 (2.6%)	0 (0%)	2 (1.2%)
Missing	68 (88.3%)	84 (87.5%)	152 (87.9%)
has.ring_f			
0	17 (22.1%)	26 (27.1%)	43 (24.9%)
1	60 (77.9%)	70 (72.9%)	130 (75.1%)
has.ring_t			
0	60 (77.9%)	70 (72.9%)	130 (75.1%)
1	17 (22.1%)	26 (27.1%)	43 (24.9%)
ring.type_e			
0	70 (90.9%)	88 (91.7%)	158 (91.3%)
1	3 (3.9%)	5 (5.2%)	8 (4.6%)
Missing	4 (5.2%)	3 (3.1%)	7 (4.0%)
ring.type_f			
0	12 (15.6%)	17 (17.7%)	29 (16.8%)
1	61 (79.2%)	76 (79.2%)	137 (79.2%)
Missing	4 (5.2%)	3 (3.1%)	7 (4.0%)
ring.type_g			
0	71 (92.2%)	90 (93.8%)	161 (93.1%)
1	2 (2.6%)	3 (3.1%)	5 (2.9%)
Missing	4 (5.2%)	3 (3.1%)	7 (4.0%)
ring.type_l			
0	69 (89.6%)	91 (94.8%)	160 (92.5%)
1	4 (5.2%)	2 (2.1%)	6 (3.5%)
Missing	4 (5.2%)	3 (3.1%)	7 (4.0%)
ring.type_m			
0	72 (93.5%)	93 (96.9%)	165 (95.4%)
1	1 (1.3%)	0 (0%)	1 (0.6%)
Missing	4 (5.2%)	3 (3.1%)	7 (4.0%)
ring.type_p			
0	71 (92.2%)	90 (93.8%)	161 (93.1%)
1	2 (2.6%)	3 (3.1%)	5 (2.9%)
Missing	4 (5.2%)	3 (3.1%)	7 (4.0%)
ring.type_r			
0	70 (90.9%)	91 (94.8%)	161 (93.1%)

(continued)

	e	p	Overall
1	3 (3.9%)	2 (2.1%)	5 (2.9%)
Missing	4 (5.2%)	3 (3.1%)	7 (4.0%)
ring.type_z			
0	73 (94.8%)	87 (90.6%)	160 (92.5%)
1	0 (0%)	6 (6.3%)	6 (3.5%)
Missing	4 (5.2%)	3 (3.1%)	7 (4.0%)
Spore.print.color_g			
0	4 (5.2%)	13 (13.5%)	17 (9.8%)
1	1 (1.3%)	0 (0%)	1 (0.6%)
Missing	72 (93.5%)	83 (86.5%)	155 (89.6%)
Spore.print.color_k			
0	4 (5.2%)	7 (7.3%)	11 (6.4%)
1	1 (1.3%)	6 (6.3%)	7 (4.0%)
Missing	72 (93.5%)	83 (86.5%)	155 (89.6%)
Spore.print.color_n			
0	5 (6.5%)	10 (10.4%)	15 (8.7%)
1	0 (0%)	3 (3.1%)	3 (1.7%)
Missing	72 (93.5%)	83 (86.5%)	155 (89.6%)
Spore.print.color_p			
0	4 (5.2%)	10 (10.4%)	14 (8.1%)
1	1 (1.3%)	3 (3.1%)	4 (2.3%)
Missing	72 (93.5%)	83 (86.5%)	155 (89.6%)
Spore.print.color_r			
0	5 (6.5%)	12 (12.5%)	17 (9.8%)
1	0 (0%)	1 (1.0%)	1 (0.6%)
Missing	72 (93.5%)	83 (86.5%)	155 (89.6%)
Spore.print.color_u			
0	5 (6.5%)	12 (12.5%)	17 (9.8%)
1	0 (0%)	1 (1.0%)	1 (0.6%)
Missing	72 (93.5%)	83 (86.5%)	155 (89.6%)
Spore.print.color_w			
0	3 (3.9%)	11 (11.5%)	14 (8.1%)
1	2 (2.6%)	2 (2.1%)	4 (2.3%)
Missing	72 (93.5%)	83 (86.5%)	155 (89.6%)
habitat_d			
0	8 (10.4%)	14 (14.6%)	22 (12.7%)
1	69 (89.6%)	82 (85.4%)	151 (87.3%)

(continued)

	e	p	Overall
habitat_g			
0	62 (80.5%)	73 (76.0%)	135 (78.0%)
1	15 (19.5%)	23 (24.0%)	38 (22.0%)
habitat_h			
0	72 (93.5%)	88 (91.7%)	160 (92.5%)
1	5 (6.5%)	8 (8.3%)	13 (7.5%)
habitat_l			
0	66 (85.7%)	89 (92.7%)	155 (89.6%)
1	11 (14.3%)	7 (7.3%)	18 (10.4%)
habitat_m			
0	69 (89.6%)	87 (90.6%)	156 (90.2%)
1	8 (10.4%)	9 (9.4%)	17 (9.8%)
habitat_p			
0	77 (100%)	94 (97.9%)	171 (98.8%)
1	0 (0%)	2 (2.1%)	2 (1.2%)
habitat_u			
0	76 (98.7%)	96 (100%)	172 (99.4%)
1	1 (1.3%)	0 (0%)	1 (0.6%)
habitat_w			
0	76 (98.7%)	96 (100%)	172 (99.4%)
1	1 (1.3%)	0 (0%)	1 (0.6%)
season_a			
0	3 (3.9%)	2 (2.1%)	5 (2.9%)
1	74 (96.1%)	94 (97.9%)	168 (97.1%)
season_s			
0	65 (84.4%)	85 (88.5%)	150 (86.7%)
1	12 (15.6%)	11 (11.5%)	23 (13.3%)
season_u			
0	16 (20.8%)	17 (17.7%)	33 (19.1%)
1	61 (79.2%)	79 (82.3%)	140 (80.9%)
season_w			
0	52 (67.5%)	80 (83.3%)	132 (76.3%)
1	25 (32.5%)	16 (16.7%)	41 (23.7%)