2306aml109-byogeshwar-assignment-5

July 1, 2023

1 Mushroom Database

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as pl
```

[2]:
```
features =␣
↪["class","cap-shape","cap-surface","cap-color","bruises","odor","gill-attachment","gill-spa

agri = pd.read_csv('agaricus-lepiota.data', names=features)
agri
```

[2]:

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	\
0	p	x	s	n	t	p	f	
1	e	x	s	y	t	a	f	
2	e	b	s	w	t	l	f	
3	p	x	y	w	t	p	f	
4	e	x	s	g	f	n	f	
…	…	…	…	…	…	…	…	
8119	e	k	s	n	f	n	a	
8120	e	x	s	n	f	n	a	
8121	e	f	s	n	f	n	a	
8122	p	k	y	n	f	y	f	
8123	e	x	s	n	f	n	a	

	gill-spacing	gill-size	gill-color	…	stalk-surface-below-ring	\
0	c	n	k	…	s	
1	c	b	k	…	s	
2	c	b	n	…	s	
3	c	n	n	…	s	
4	w	b	k	…	s	
…	…	…	…	…	…	
8119	c	b	y	…	s	
8120	c	b	y	…	s	
8121	c	b	n	…	s	
8122	c	n	b	…	k	

```
8123            c         b          y  …                        s

     stalk-color-above-ring stalk-color-below-ring veil-type veil-color  \
0                         w                      w         p          w
1                         w                      w         p          w
2                         w                      w         p          w
3                         w                      w         p          w
4                         w                      w         p          w
…                       …                      …       …        …
8119                      o                      o         p          o
8120                      o                      o         p          n
8121                      o                      o         p          o
8122                      w                      w         p          w
8123                      o                      o         p          o

     ring-number ring-type spore-print-color population habitat
0              o         p                 k          s       u
1              o         p                 n          n       g
2              o         p                 n          n       m
3              o         p                 k          s       u
4              o         e                 n          a       g
…            …       …                 …        …     …
8119           o         p                 b          c       l
8120           o         p                 b          v       l
8121           o         p                 b          c       l
8122           o         e                 w          v       l
8123           o         p                 o          c       l

[8124 rows x 23 columns]
```

[3]:
```
agri.shape
```

[3]:
```
(8124, 23)
```

Basic Information of Data type:

[4]:
```
agri.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   class                     8124 non-null   object
 1   cap-shape                 8124 non-null   object
 2   cap-surface               8124 non-null   object
 3   cap-color                 8124 non-null   object
 4   bruises                   8124 non-null   object
```

```
 5   odor                      8124 non-null   object
 6   gill-attachment           8124 non-null   object
 7   gill-spacing              8124 non-null   object
 8   gill-size                 8124 non-null   object
 9   gill-color                8124 non-null   object
 10  stalk-shape               8124 non-null   object
 11  stalk-root                8124 non-null   object
 12  stalk-surface-above-ring  8124 non-null   object
 13  stalk-surface-below-ring  8124 non-null   object
 14  stalk-color-above-ring    8124 non-null   object
 15  stalk-color-below-ring    8124 non-null   object
 16  veil-type                 8124 non-null   object
 17  veil-color                8124 non-null   object
 18  ring-number               8124 non-null   object
 19  ring-type                 8124 non-null   object
 20  spore-print-color         8124 non-null   object
 21  population                8124 non-null   object
 22  habitat                   8124 non-null   object
dtypes: object(23)
memory usage: 1.4+ MB
```

Overview of all attributes at dataset:

[5]:
```
agri.describe()
```

[5]:

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	…
count	8124	8124	8124	8124	8124	8124	8124	…
unique	2	6	4	10	2	9	2	…
top	e	x	y	n	f	n	f	…
freq	4208	3656	3244	2284	4748	3528	7914	…

	gill-spacing	gill-size	gill-color	…	stalk-surface-below-ring
count	8124	8124	8124	…	8124
unique	2	2	12	…	4
top	c	b	b	…	s
freq	6812	5612	1728	…	4936

	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color
count	8124	8124	8124	8124
unique	9	9	1	4
top	w	w	p	w
freq	4464	4384	8124	7924

	ring-number	ring-type	spore-print-color	population	habitat
count	8124	8124	8124	8124	8124
unique	3	5	9	6	7
top	o	p	w	v	d
freq	7488	3968	2388	4040	3148

```
[4 rows x 23 columns]
```

[6]:
```
agri["class"].value_counts()
```

[6]:
```
e    4208
p    3916
Name: class, dtype: int64
```

[7]:
```
agri.isnull().sum()
```

[7]:
```
class                       0
cap-shape                   0
cap-surface                 0
cap-color                   0
bruises                     0
odor                        0
gill-attachment             0
gill-spacing                0
gill-size                   0
gill-color                  0
stalk-shape                 0
stalk-root                  0
stalk-surface-above-ring    0
stalk-surface-below-ring    0
stalk-color-above-ring      0
stalk-color-below-ring      0
veil-type                   0
veil-color                  0
ring-number                 0
ring-type                   0
spore-print-color           0
population                  0
habitat                     0
dtype: int64
```

[8]:
```
agri["class"].unique()
```

[8]:
```
array(['p', 'e'], dtype=object)
```

Analysis relation between CLASS & ODOR

[9]:
```
pd.crosstab(agri["class"], agri["odor"],margins=True)
```

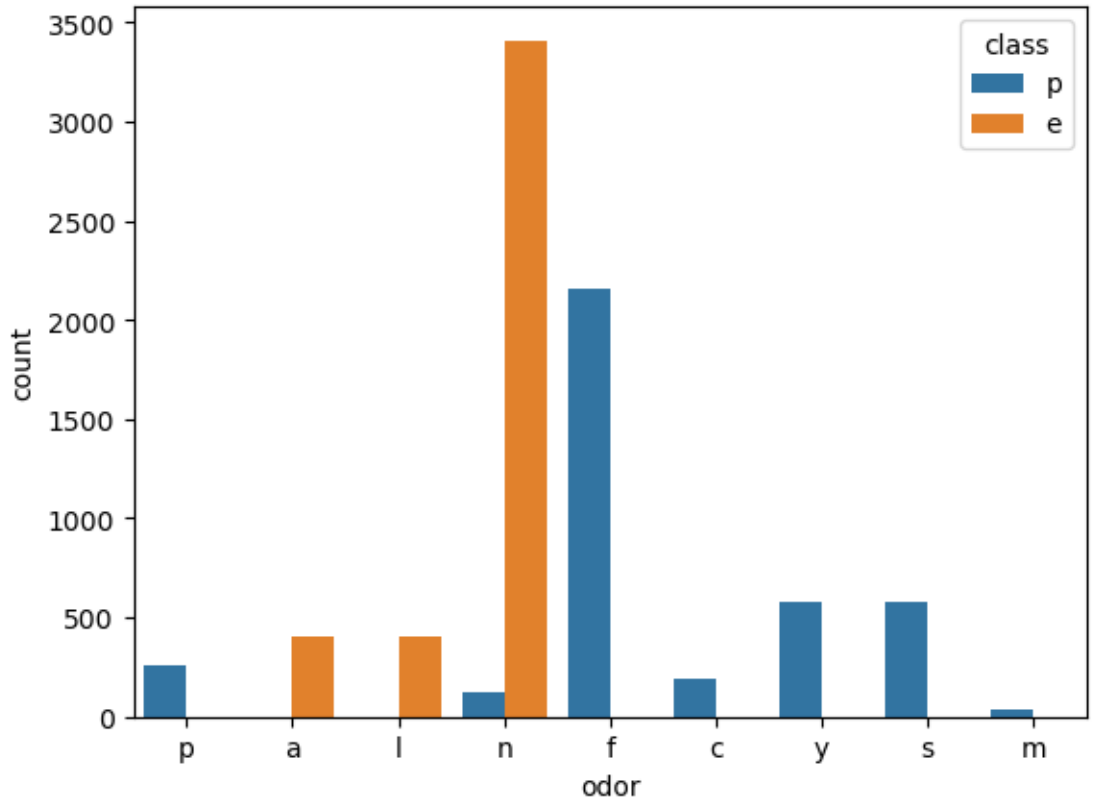
[9]:

odor	a	c	f	l	m	n	p	s	y	All
class										
e	400	0	0	400	0	3408	0	0	0	4208

```
p          0  192  2160    0  36   120  256  576  576  3916
All      400  192  2160  400  36  3528  256  576  576  8124
```

[10]:
```
sns.countplot(data=agri, x="odor", hue="class")
```

[10]:
```
<Axes: xlabel='odor', ylabel='count'>
```



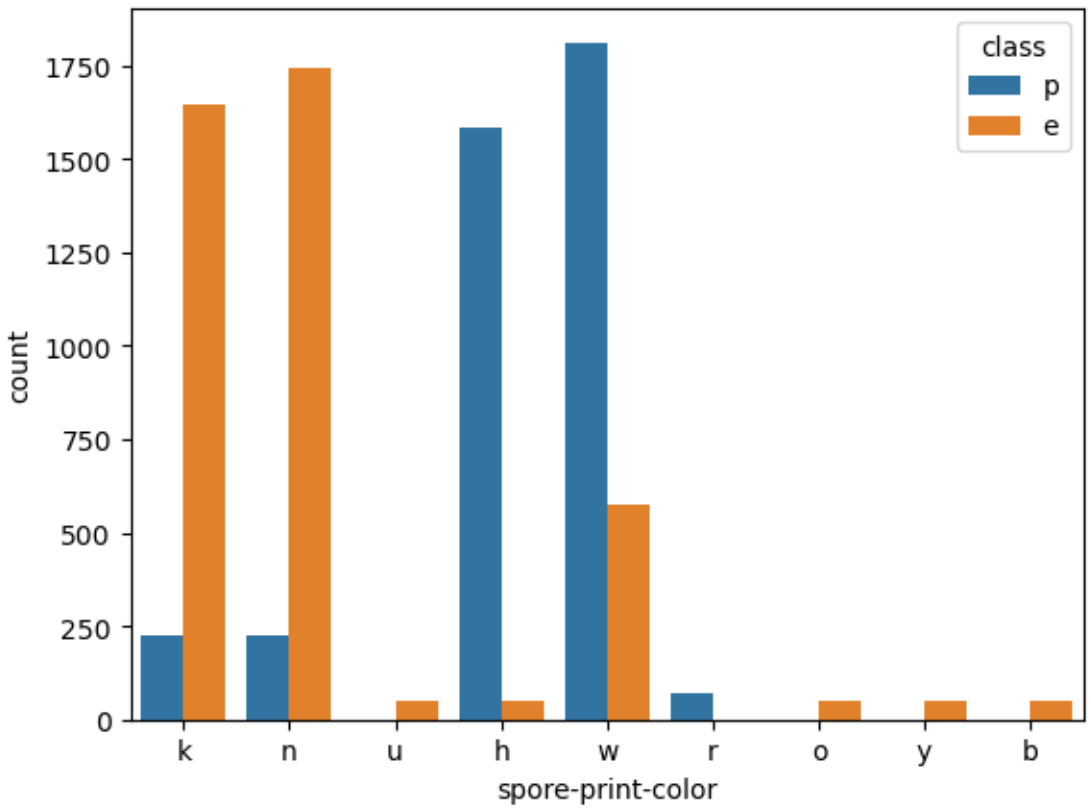
Analysis relation between CLASS & SPORE PRINT COLOR

[11]:
```
pd.crosstab(agri["class"], agri["spore-print-color"],margins=True)
```

[11]:
```
spore-print-color    b     h     k     n   o   r   u     w   y   All
class
e                   48    48  1648  1744  48   0  48   576  48  4208
p                    0  1584   224   224   0  72   0  1812   0  3916
All                 48  1632  1872  1968  48  72  48  2388  48  8124
```

[12]:
```
sns.countplot(data=agri, x="spore-print-color", hue="class")
```

[12]:
```
<Axes: xlabel='spore-print-color', ylabel='count'>
```



Analysis relation between CLASS & POPULATION

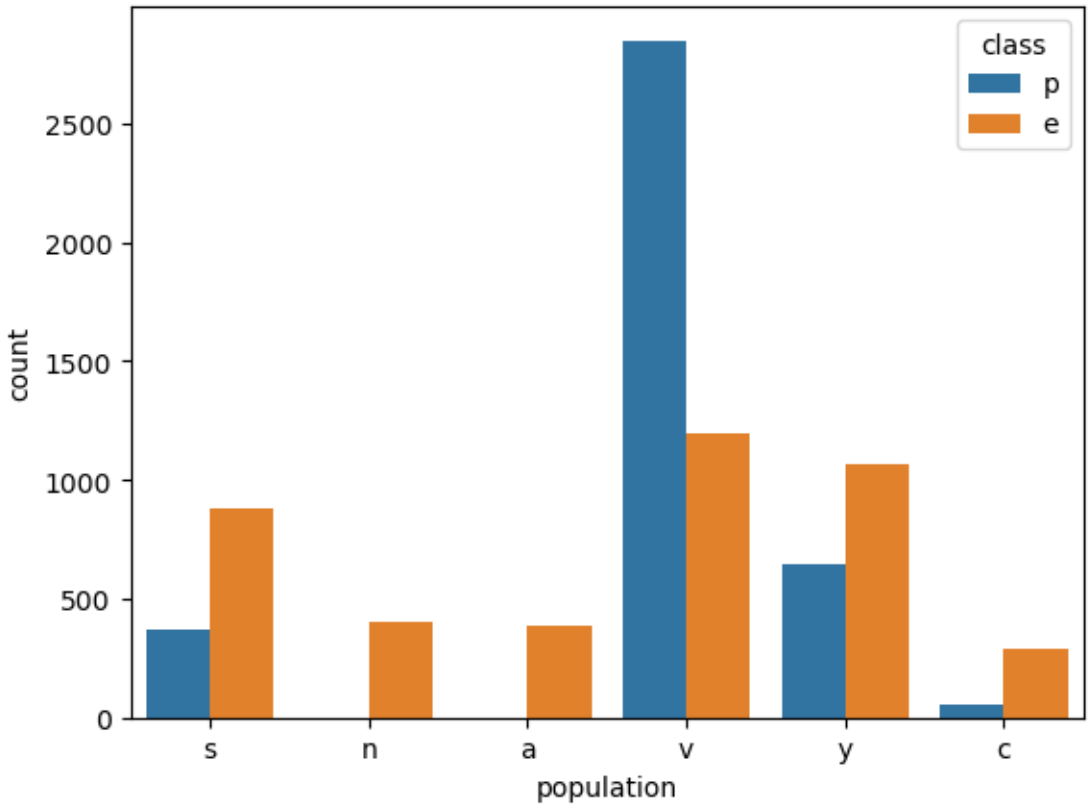
[13]:
```
pd.crosstab(agri["class"], agri["population"],margins=True)
```

[13]:

population	a	c	n	s	v	y	All
class							
e	384	288	400	880	1192	1064	4208
p	0	52	0	368	2848	648	3916
All	384	340	400	1248	4040	1712	8124

[14]:
```
sns.countplot(data=agri, x="population", hue="class")
```

[14]:
```
<Axes: xlabel='population', ylabel='count'>
```



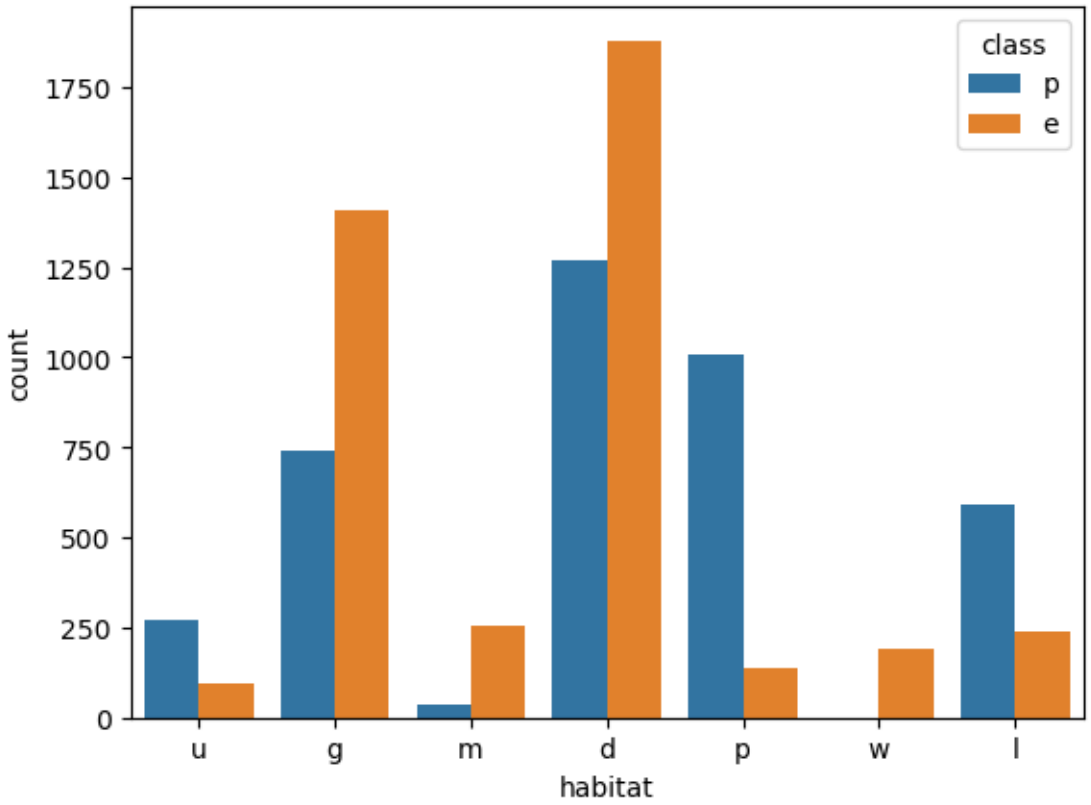
Analysis relation between CLASS & HABITAT

[15]:
```
pd.crosstab(agri["class"], agri["habitat"],margins=True)
```

[15]:
```
habitat     d     g    l    m     p    u    w   All
class
e        1880  1408  240  256   136   96  192  4208
p        1268   740  592   36  1008  272    0  3916
All      3148  2148  832  292  1144  368  192  8124
```

[16]:
```
sns.countplot(data=agri, x="habitat", hue="class")
```

[16]:
```
<Axes: xlabel='habitat', ylabel='count'>
```



Dropping Veil-type attribute as unique categorical values is '1' only.

[17]:
```
agri1 = agri.drop(["veil-type"], axis = 1)
agri1
```

[17]:

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	\
0	p	x	s	n	t	p	f	
1	e	x	s	y	t	a	f	
2	e	b	s	w	t	l	f	
3	p	x	y	w	t	p	f	
4	e	x	s	g	f	n	f	
…	…	…	…	…	…	…	…	
8119	e	k	s	n	f	n	a	
8120	e	x	s	n	f	n	a	
8121	e	f	s	n	f	n	a	
8122	p	k	y	n	f	y	f	
8123	e	x	s	n	f	n	a	

	gill-spacing	gill-size	gill-color	…	stalk-surface-above-ring	\
0	c	n	k	…	s	
1	c	b	k	…	s	

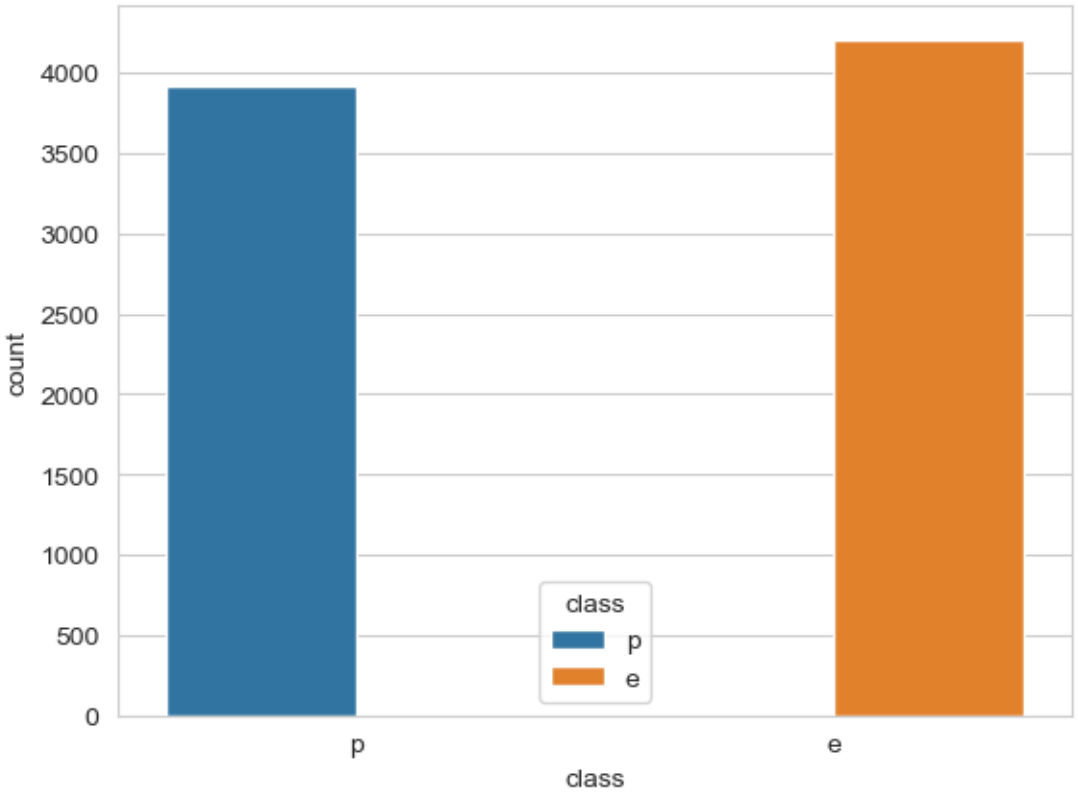
```
2              c         b          n  …                         s
3              c         n          n  …                         s
4              w         b          k  …                         s
…            …       …          …  …                       …
8119           c         b          y  …                         s
8120           c         b          y  …                         s
8121           c         b          n  …                         s
8122           c         n          b  …                         s
8123           c         b          y  …                         s

     stalk-surface-below-ring stalk-color-above-ring stalk-color-below-ring  \
0                           s                      w                      w
1                           s                      w                      w
2                           s                      w                      w
3                           s                      w                      w
4                           s                      w                      w
…                         …                      …                      …
8119                        s                      o                      o
8120                        s                      o                      o
8121                        s                      o                      o
8122                        k                      w                      w
8123                        s                      o                      o

     veil-color ring-number ring-type spore-print-color population habitat
0             w           o         p                 k          s       u
1             w           o         p                 n          n       g
2             w           o         p                 n          n       m
3             w           o         p                 k          s       u
4             w           o         e                 n          a       g
…           …         …       …               …        …     …
8119          o           o         p                 b          c       l
8120          n           o         p                 b          v       l
8121          o           o         p                 b          c       l
8122          w           o         e                 w          v       l
8123          o           o         p                 o          c       l

[8124 rows x 22 columns]
```

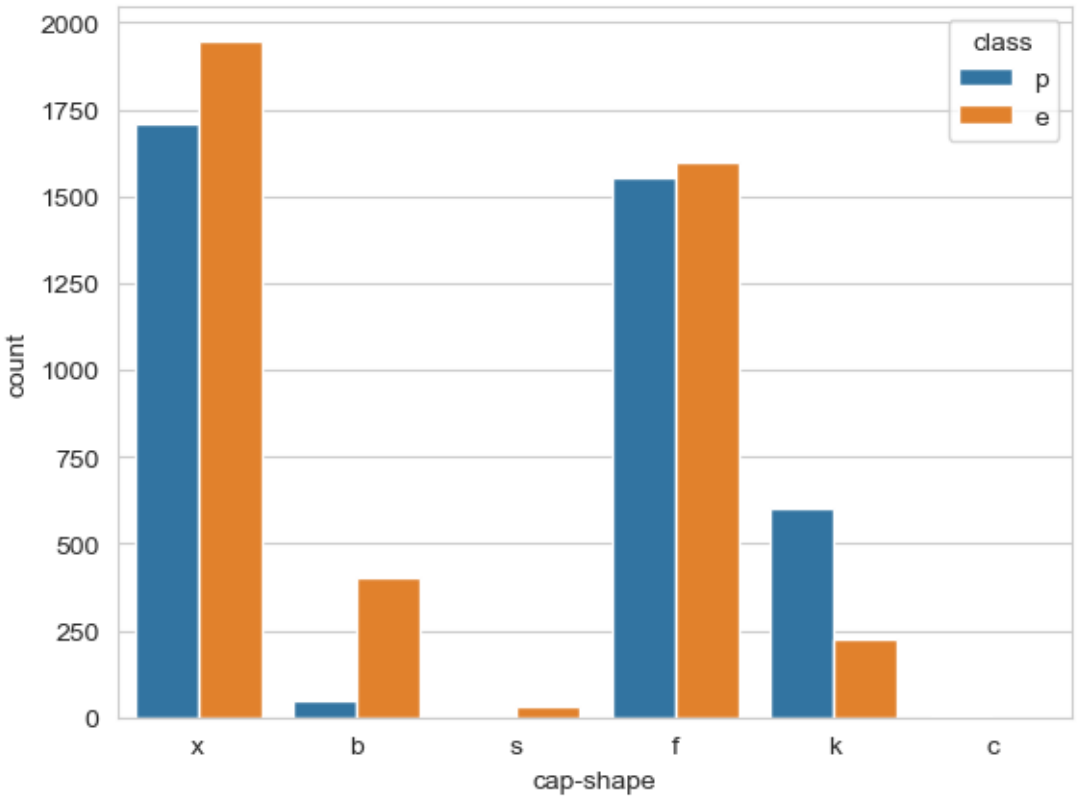
Generating countplot with each feature comparing with Class

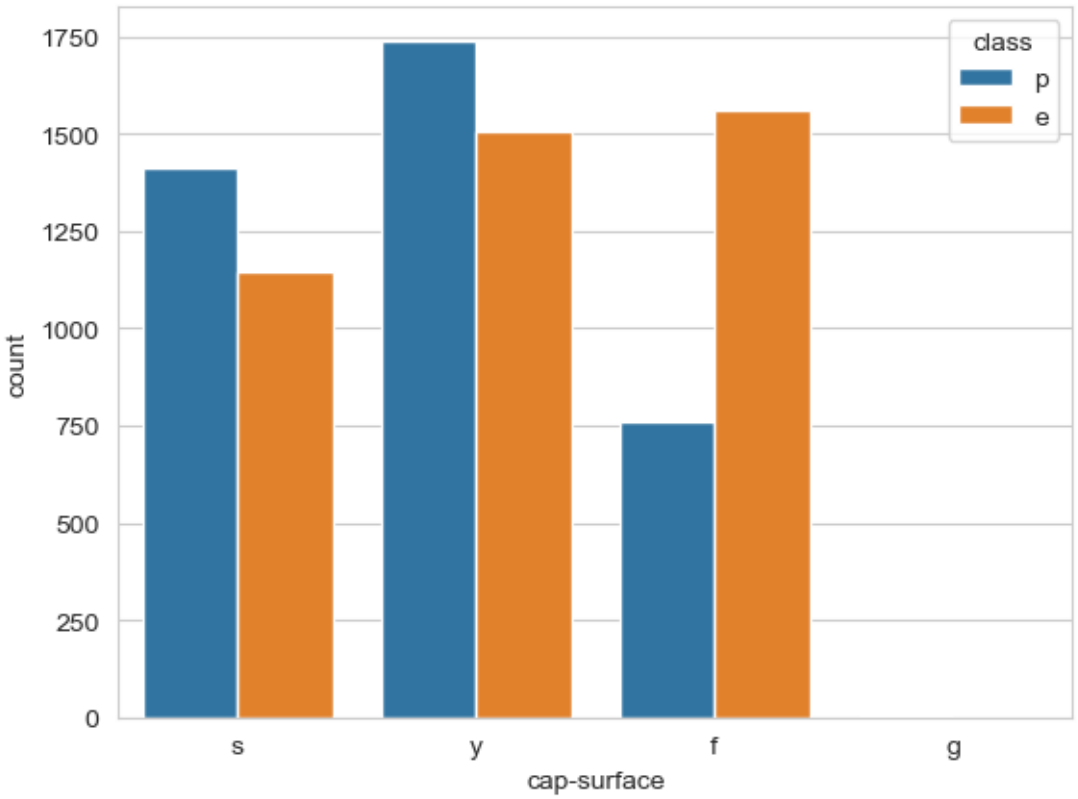
[18]:
```
sns.set_style('whitegrid')
for i, col in enumerate(agri1.columns):
    pl.figure(i)
    sns_plot = sns.countplot(x=col, hue='class', data=agri)
    sns_plot.figure.savefig("{} countplot.png".format(col))
```

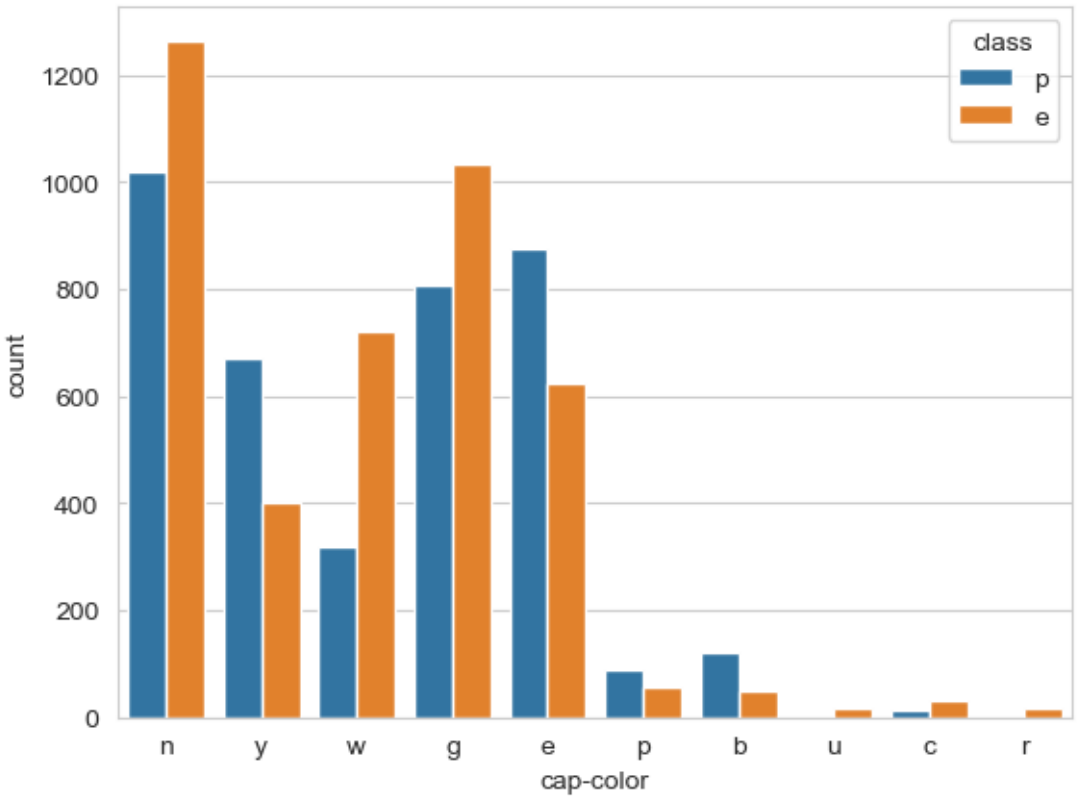
```
C:\Users\Lenovo\AppData\Local\Temp\ipykernel_4168\2902536500.py:3:
RuntimeWarning: More than 20 figures have been opened. Figures created through
```

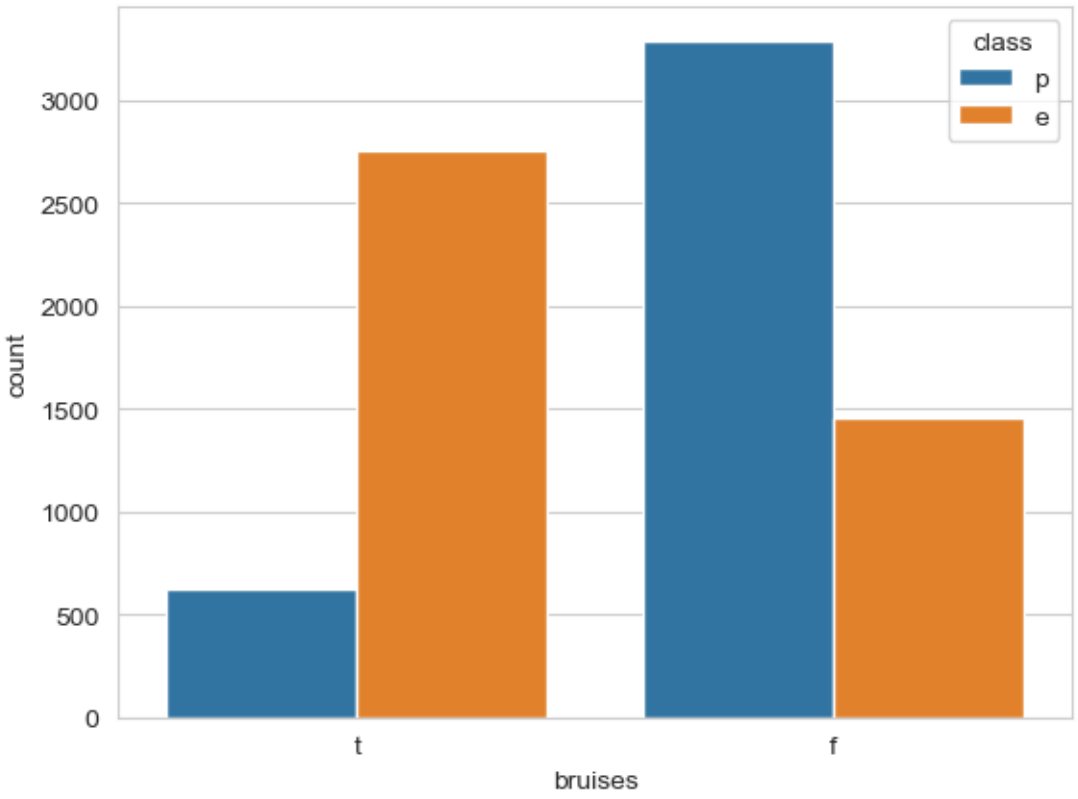
```
the pyplot interface (`matplotlib.pyplot.figure`) are retained until explicitly
closed and may consume too much memory. (To control this warning, see the
rcParam `figure.max_open_warning`). Consider using `matplotlib.pyplot.close()`.
  pl.figure(i)
```

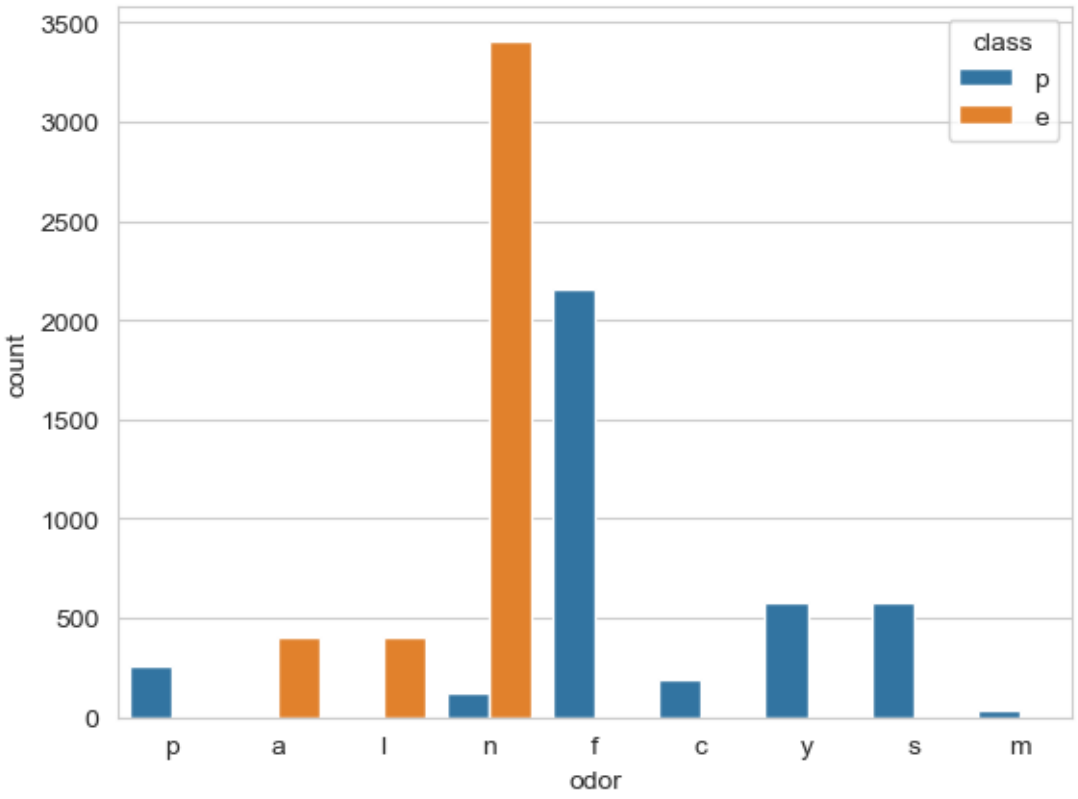


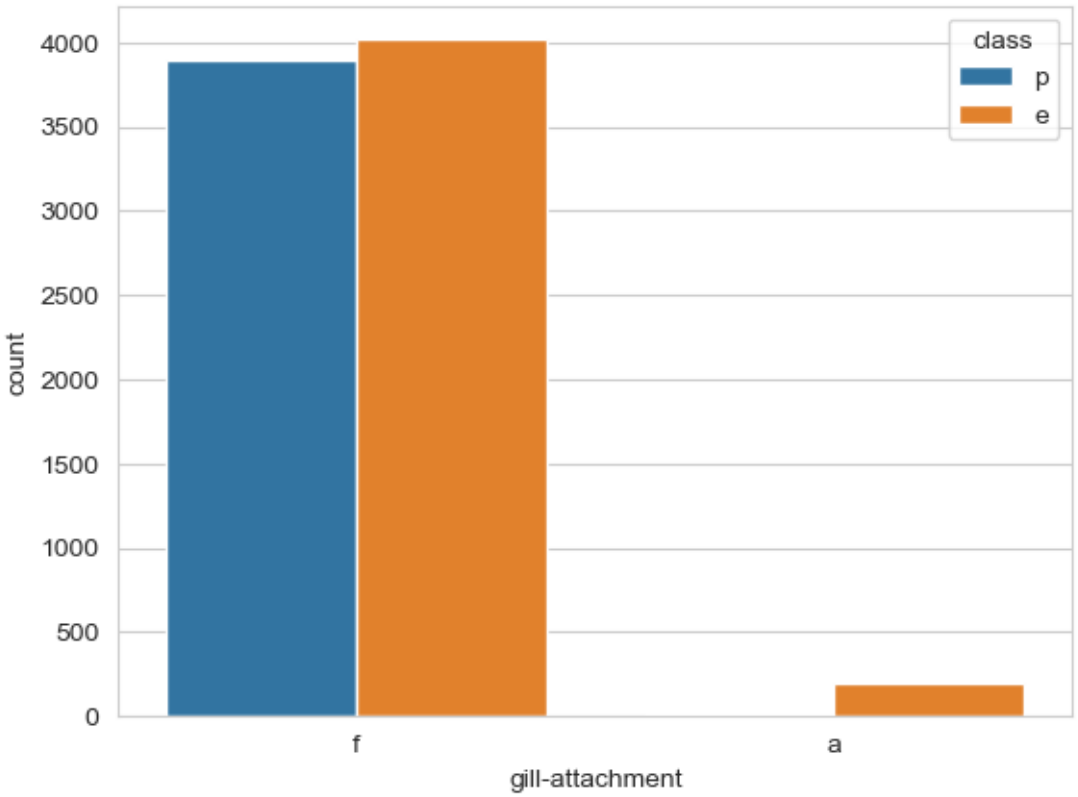

class
p
e
2000
1750
1500
1250
1000
750
500
250
0
count
x
b
s
f
k
c
cap-shape

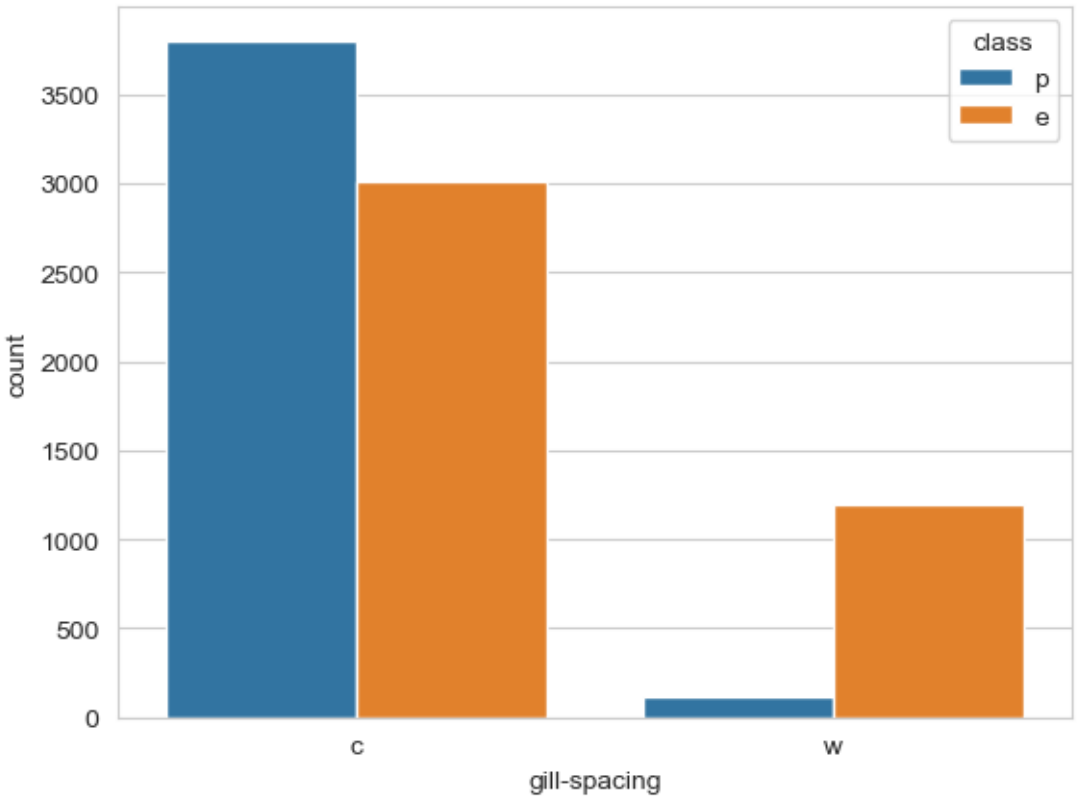

class
p
e
1750
1500
1250
1000
750
500
250
0
count
s
y
f
g
cap-surface

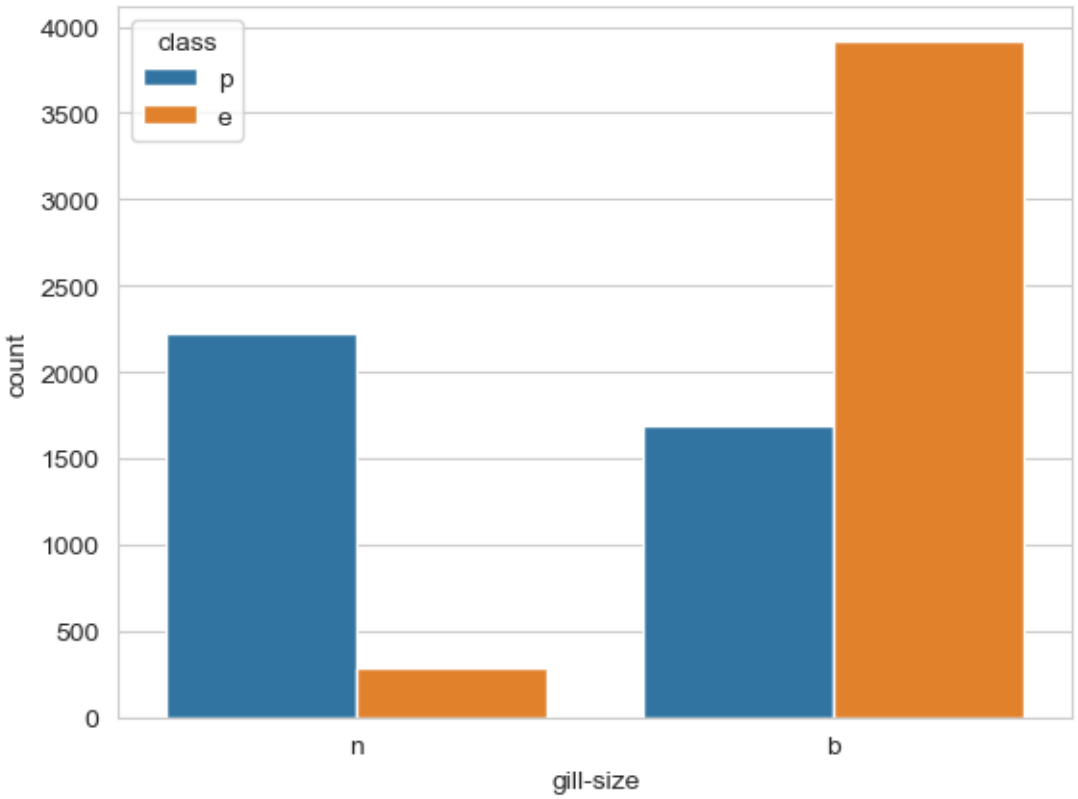

class
p
e
count
1200
1000
800
600
400
200
0
n
y
w
g
e
p
b
u
c
r
cap-color

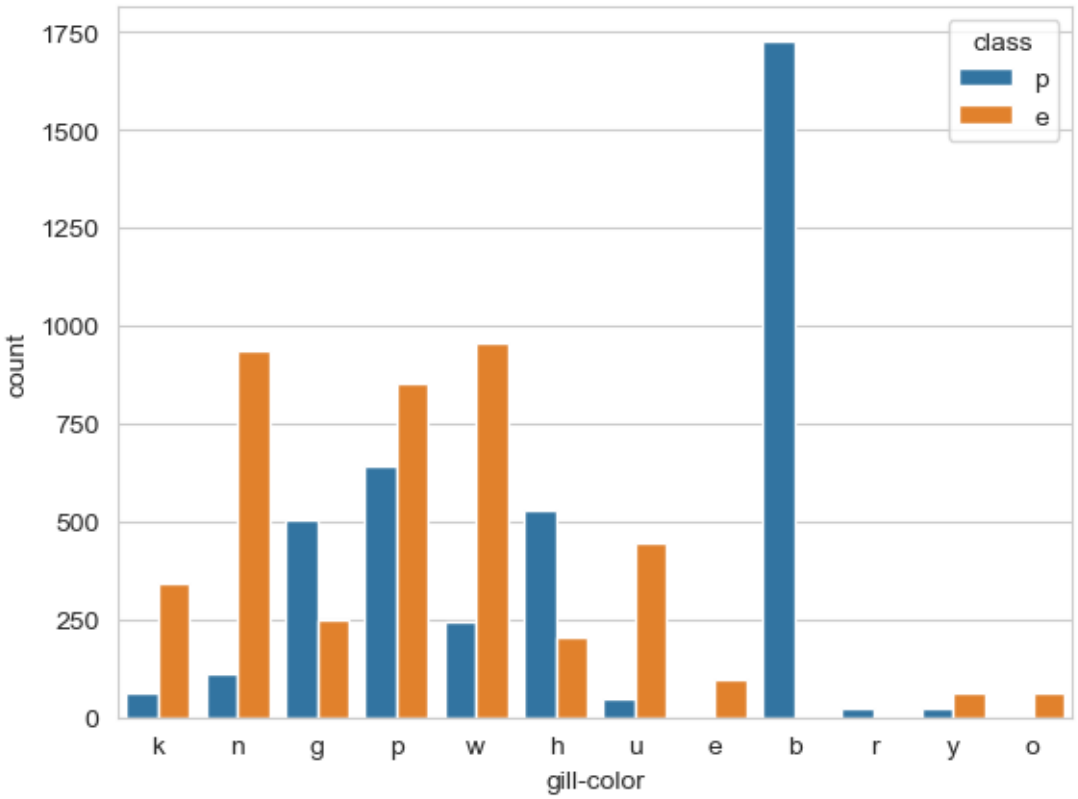

class
p
e
3000
2500
2000
1500
1000
500
0
count
t
f
bruises

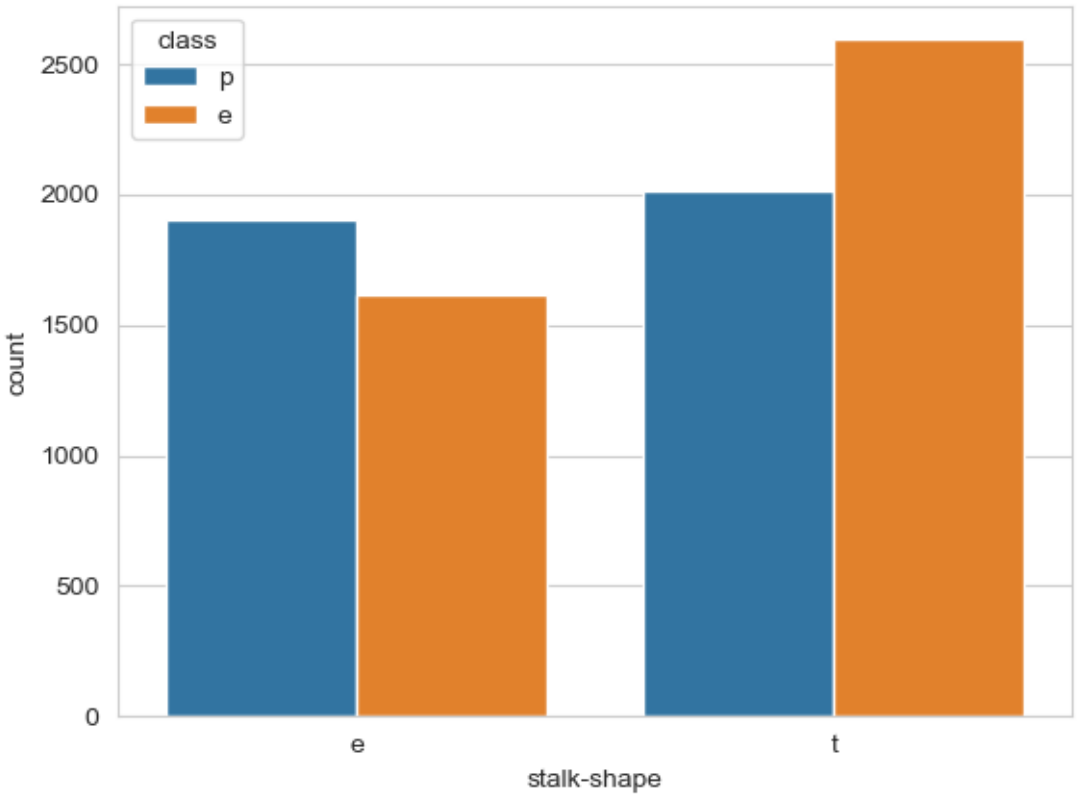

class
p
e
count
3500
3000
2500
2000
1500
1000
500
0
p
a
l
n
f
c
y
s
m
odor

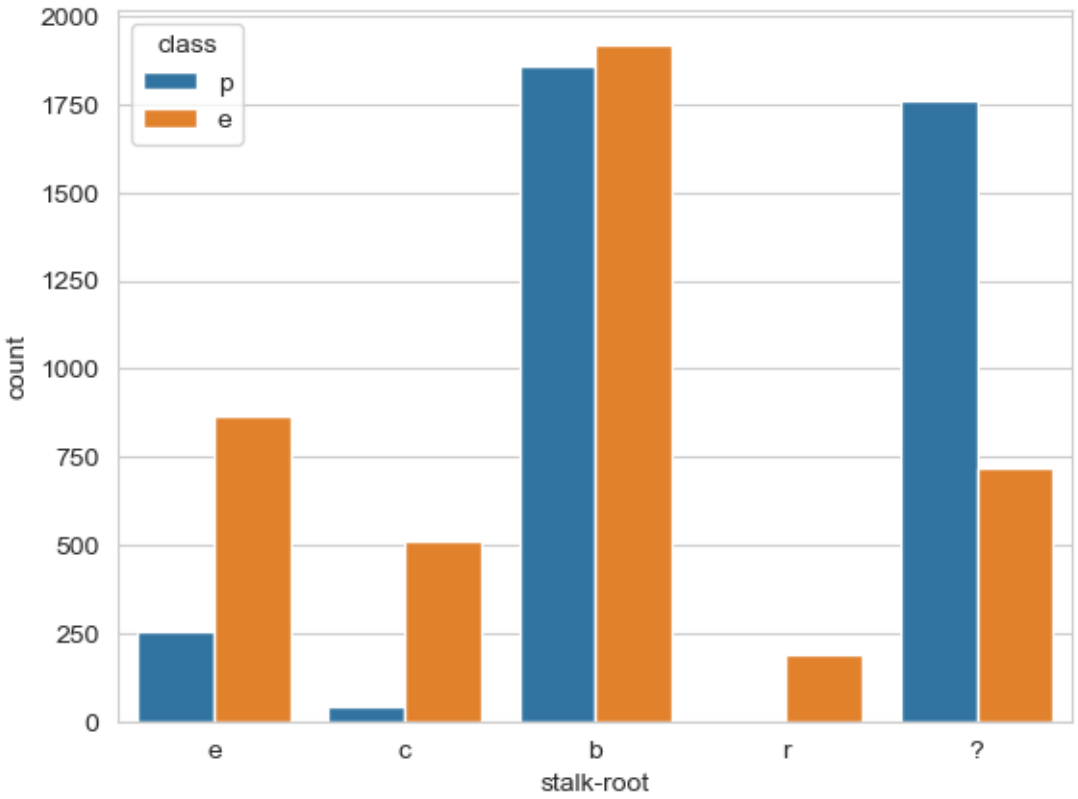

class
p
e
4000
3500
3000
2500
2000
1500
1000
500
0
count
f
a
gill-attachment

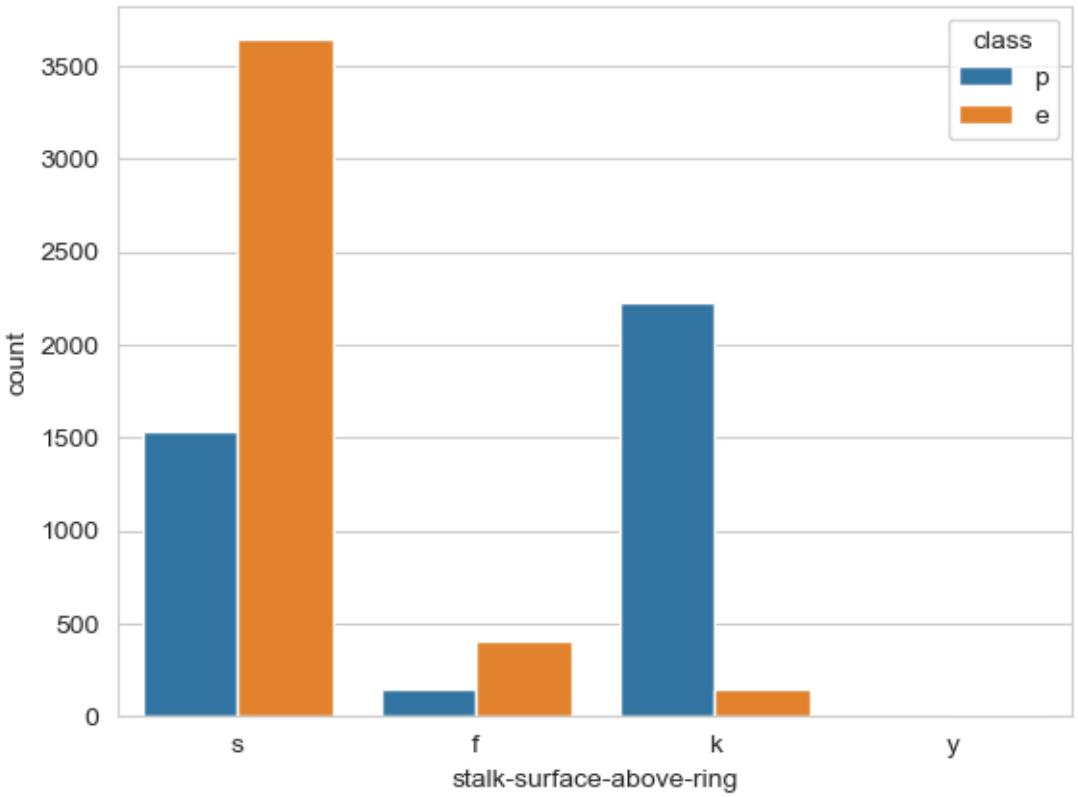

class
p
e
3500
3000
2500
2000
1500
1000
500
0
count
c
w
gill-spacing

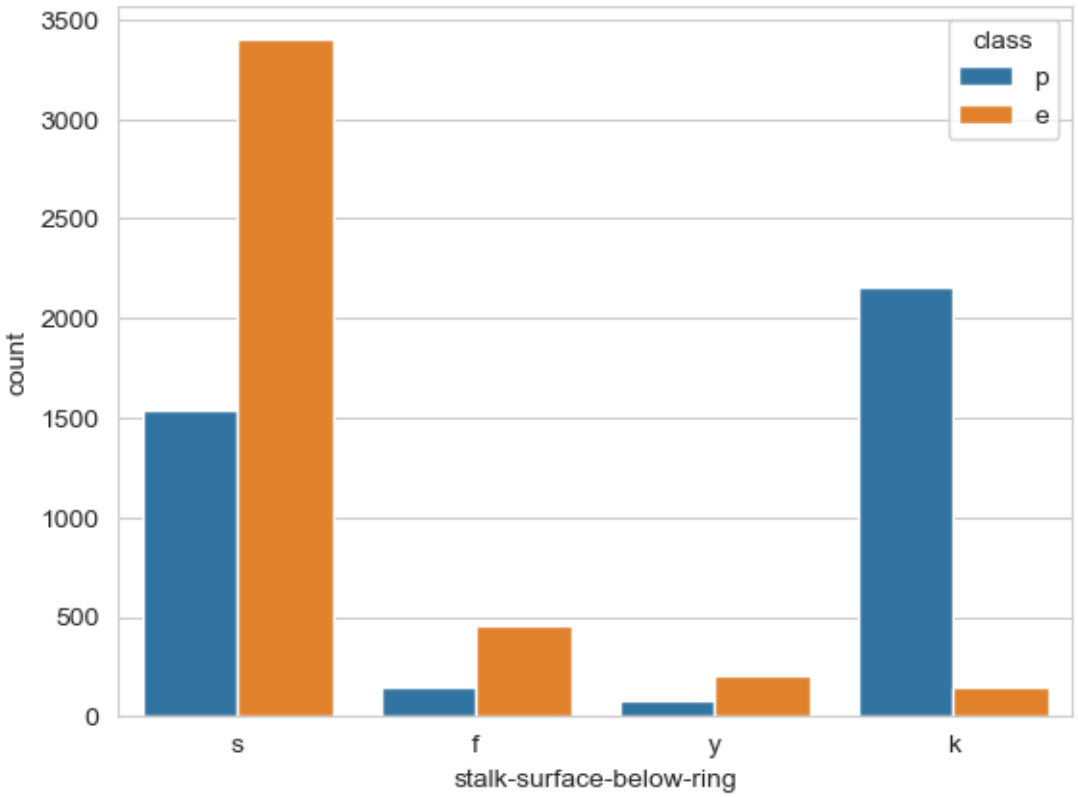

class
p
e
4000
3500
3000
2500
2000
1500
1000
500
0
count
n
b
gill-size

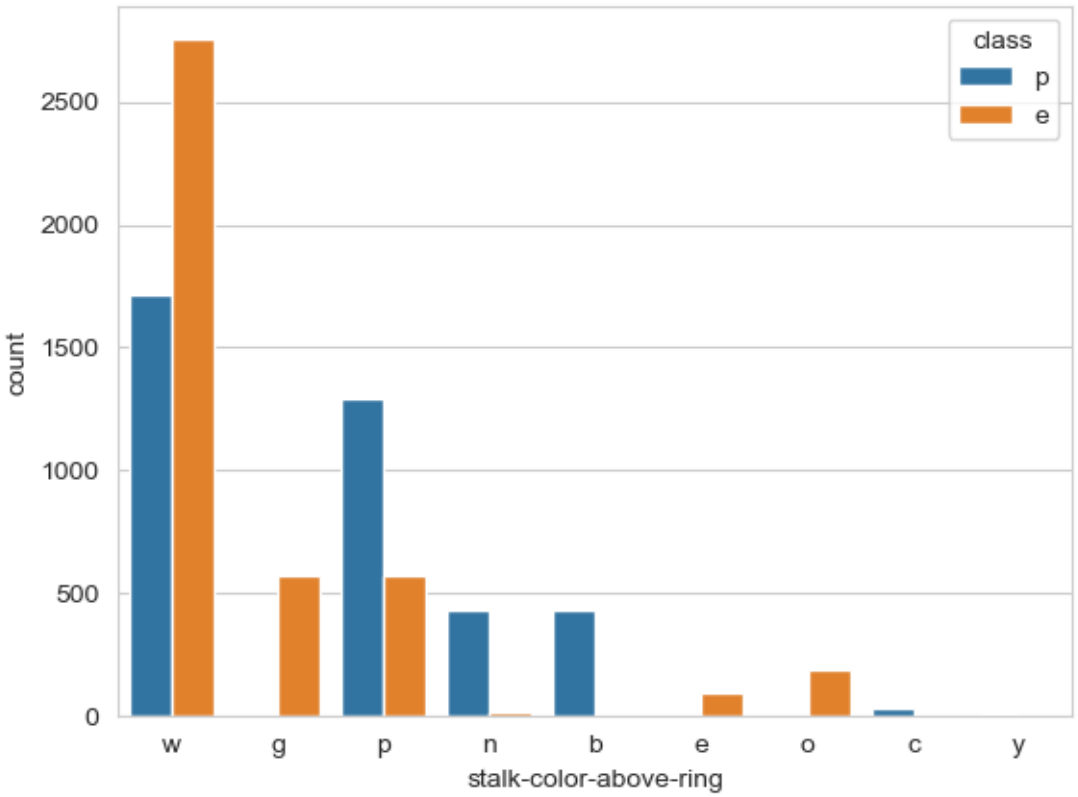

class
p
e
count
1750
1500
1250
1000
750
500
250
0
k
n
g
p
w
h
u
e
b
r
y
o
gill-color

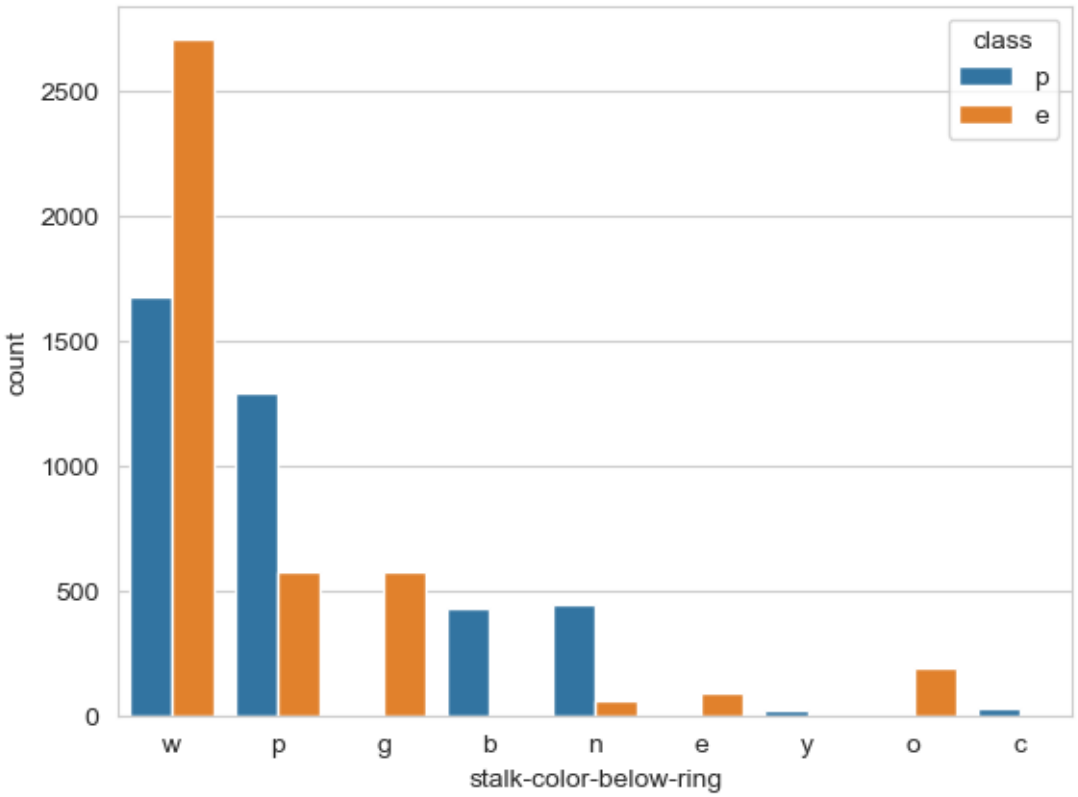

class
p
e
2500
2000
1500
1000
500
0
count
e
t
stalk-shape

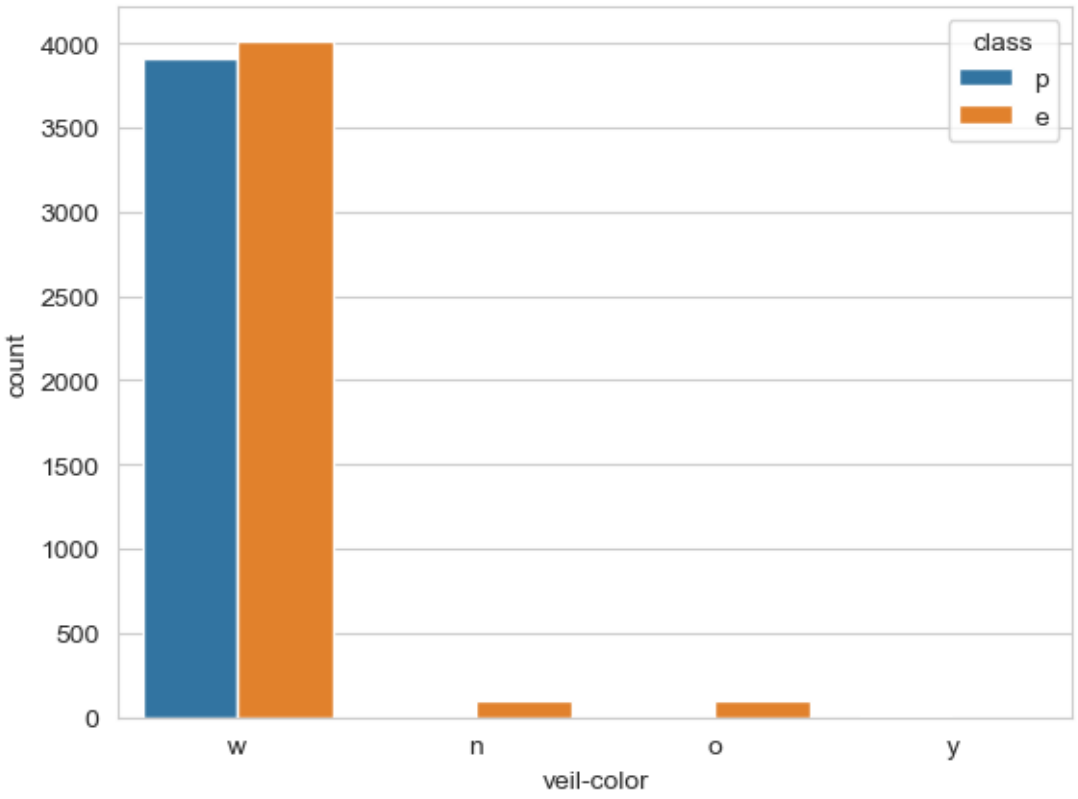

class
p
e
2000
1750
1500
1250
1000
750
500
250
0
count
e
c
b
r
?
stalk-root

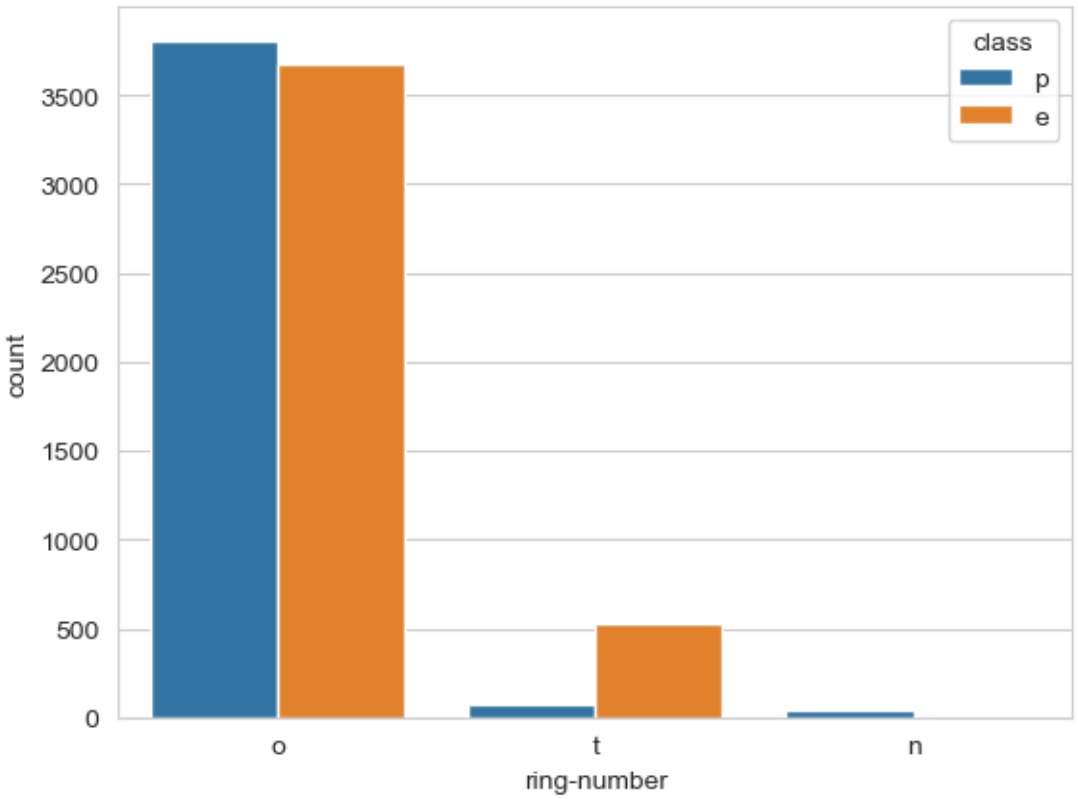

class
p
e
3500
3000
2500
2000
1500
1000
500
0
count
s
f
k
y
stalk-surface-above-ring

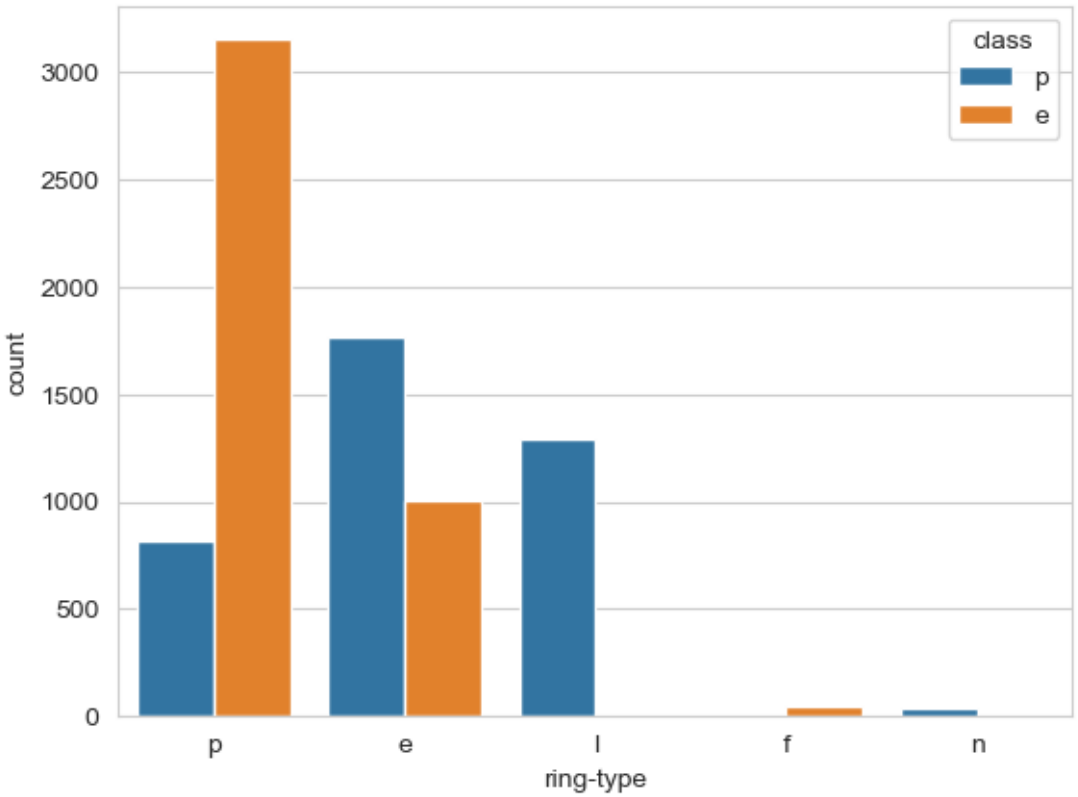

class
p
e
3500
3000
2500
2000
1500
1000
500
0
count
s
f
y
k
stalk-surface-below-ring

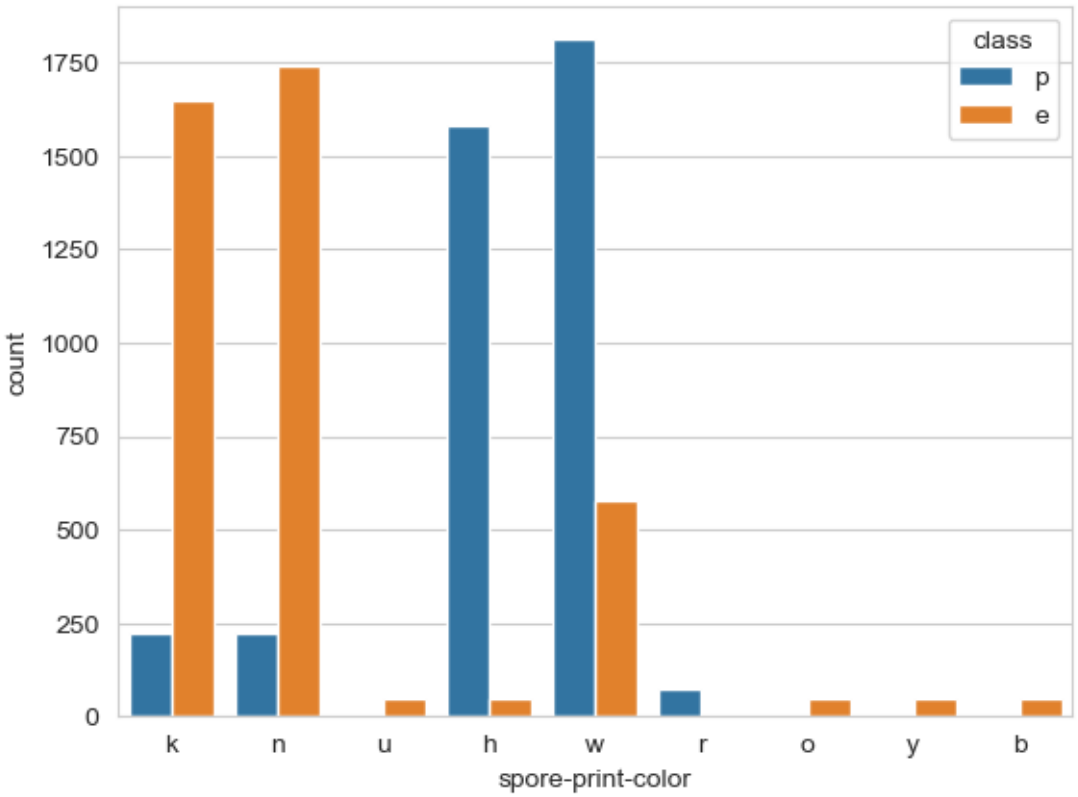

class
p
e
2500
2000
1500
1000
500
0
count
w
g
p
n
b
e
o
c
y
stalk-color-above-ring

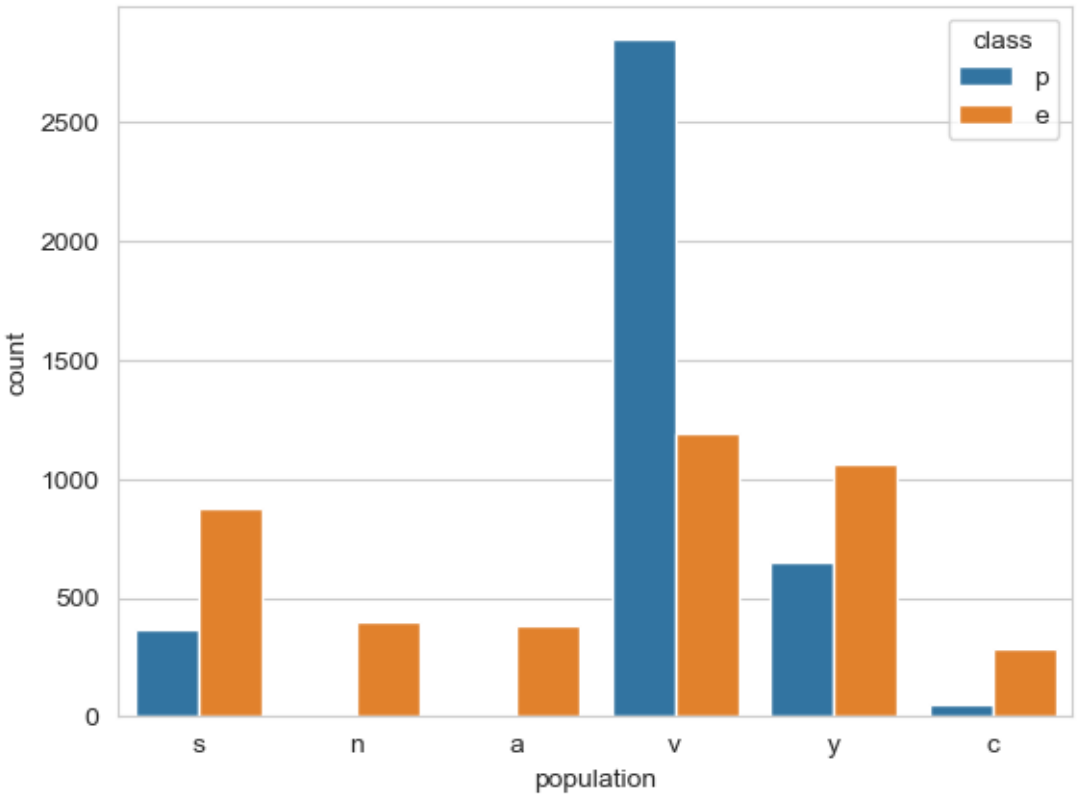

class
p
e
2500
2000
1500
1000
500
0
count
w
p
g
b
n
e
y
o
c
stalk-color-below-ring

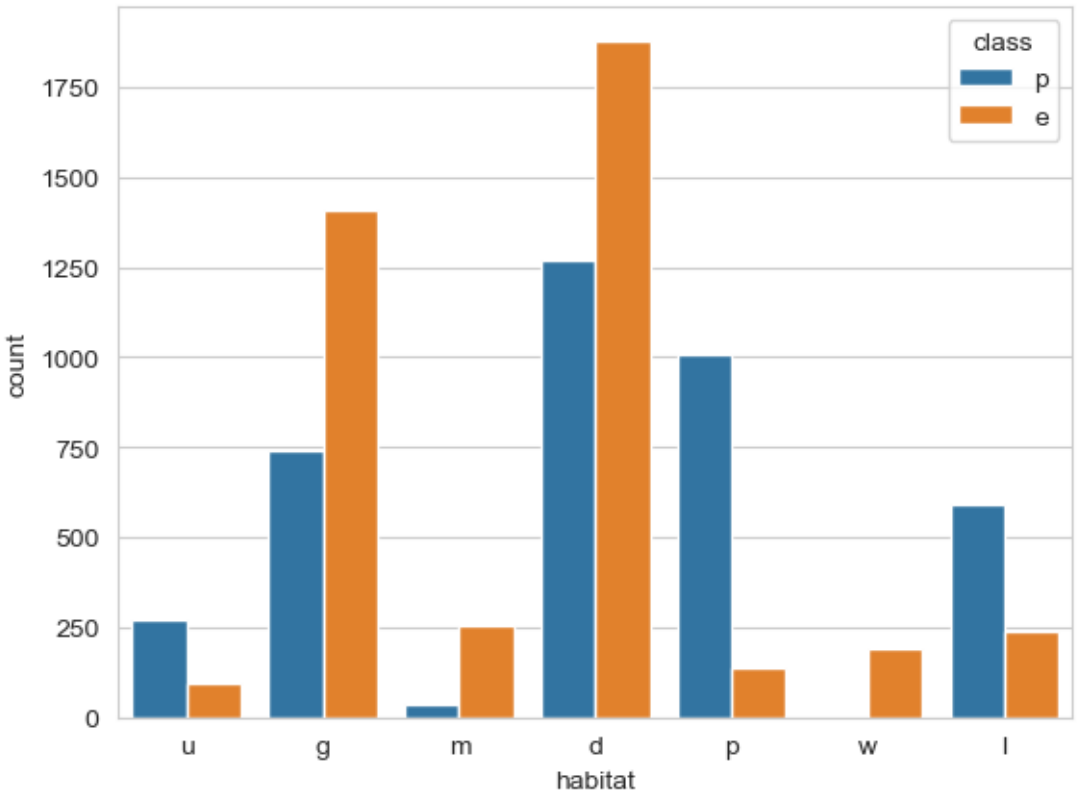

class
p
e
4000
3500
3000
2500
2000
1500
1000
500
0
count
w
n
o
y
veil-color


class
p
e
3500
3000
2500
2000
1500
1000
500
0
count
o
t
n
ring-number


class
p
e
3000
2500
2000
1500
1000
500
0
count
p
e
l
f
n
ring-type


class
p
e
1750
1500
1250
1000
750
500
250
0
count
k
n
u
h
w
r
o
y
b
spore-print-color


class
p
e
2500
2000
1500
1000
500
0
count
s
n
a
v
y
c
population


class
p
e
1750
1500
1250
1000
750
500
250
0
count
u
g
m
d
p
w
l
habitat