In [102…

```
import pandas as pd
import numpy as np
from scipy.stats import chi2_contingency
from sklearn.preprocessing import LabelEncoder

import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

import warnings
warnings.filterwarnings("ignore")
```

In [103…

```
features= ['classes','cap-shape','cap-surface','cap-color','bruises','odor','gill-attachment','gill-spacing','gill-size','gill-color','stalk-shape',
           'stalk-root','stalk-surface-above-ring','stalk-surface-below-ring','stalk-color-above-ring','stalk-color-below-ring','veil-type',
           'veil-color','ring-number','ring-type','spore-print-color','population','habitat']

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

In [104…

```
df
```

Out[104]:

	classes	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	population	habitat
0	p	x	s	n	t	p	f	c	n	k	...	s	w	w	p	w	o	p	k	s	u
1	e	x	s	y	t	a	f	c	b	k	...	s	w	w	p	w	o	p	n	n	g
2	e	b	s	w	t	l	f	c	b	n	...	s	w	w	p	w	o	p	n	n	m
3	p	x	y	w	t	p	f	c	n	n	...	s	w	w	p	w	o	p	k	s	u
4	e	x	s	g	f	n	f	w	b	k	...	s	w	w	p	w	o	e	n	a	g
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8119	e	k	s	n	f	n	a	c	b	y	...	s	o	o	p	o	o	p	b	c	l
8120	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	n	o	p	b	v	l
8121	e	f	s	n	f	n	a	c	b	n	...	s	o	o	p	o	o	p	b	c	l
8122	p	k	y	n	f	y	f	c	n	b	...	k	w	w	p	w	o	e	w	v	l
8123	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	o	o	p	o	c	l

8124 rows × 23 columns

In [105…

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   classes                   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
```

In [106…

```
df.describe()
```

Out[106]:

	classes	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	population	habitat
count	8124	8124	8124	8124	8124	8124	8124	8124	8124	8124	...	8124	8124	8124	8124	8124	8124	8124	8124	8124	8124
unique	2	6	4	10	2	9	2	2	2	12	...	4	9	9	1	4	3	5	9	6	7
top	e	x	y	n	f	n	f	c	b	b	...	s	w	w	p	w	o	p	w	v	d
freq	4208	3656	3244	2284	4748	3528	7914	6812	5612	1728	...	4936	4464	4384	8124	7924	7488	3968	2388	4040	3148

4 rows × 23 columns

In [107…

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   classes                   8124 non-null   int32
 1   cap-shape                 8124 non-null   int32
 2   cap-surface               8124 non-null   int32
 3   cap-color                 8124 non-null   int32
 4   bruises                   8124 non-null   int32
 5   odor                      8124 non-null   int32
 6   gill-attachment           8124 non-null   int32
 7   gill-spacing              8124 non-null   int32
 8   gill-size                 8124 non-null   int32
 9   gill-color                8124 non-null   int32
 10  stalk-shape               8124 non-null   int32
 11  stalk-root                8124 non-null   int32
 12  stalk-surface-above-ring  8124 non-null   int32
 13  stalk-surface-below-ring  8124 non-null   int32
 14  stalk-color-above-ring    8124 non-null   int32
 15  stalk-color-below-ring    8124 non-null   int32
 16  veil-type                 8124 non-null   int32
 17  veil-color                8124 non-null   int32
 18  ring-number               8124 non-null   int32
 19  ring-type                 8124 non-null   int32
 20  spore-print-color         8124 non-null   int32
 21  population                8124 non-null   int32
 22  habitat                   8124 non-null   int32
dtypes: int32(23)
memory usage: 730.0 KB
```

In [108…

```
contingency_table = pd.crosstab(df['classes'],df['cap-shape'])
print(contingency_table)
```

```
cap-shape    b  c     f    k   s     x
classes
e          404  0  1596  228  32  1948
p           48  4  1556  600   0  1708
```

In [109…

```
contingency_table = pd.crosstab(df['classes'], df['ring-type'])
print(contingency_table)
```

```
ring-type     e   f     l   n     p
classes
e          1008  48     0   0  3152
p          1768   0  1296  36   816
```

In [111…

```
contingency_table = pd.crosstab(df['classes'],df['population'])
chi2, p_value, _, _ = chi2_contingency(contingency_table)
print(f"Chi-square value: {chi2}")
```

```
Chi-square value: 1929.7408909028088
```

In [112…

```
contingency_table = pd.crosstab(df['classes'],df['odor'])
chi2, p_value, _, _ = chi2_contingency(contingency_table)
print(f"Chi-square value: {chi2}")
```

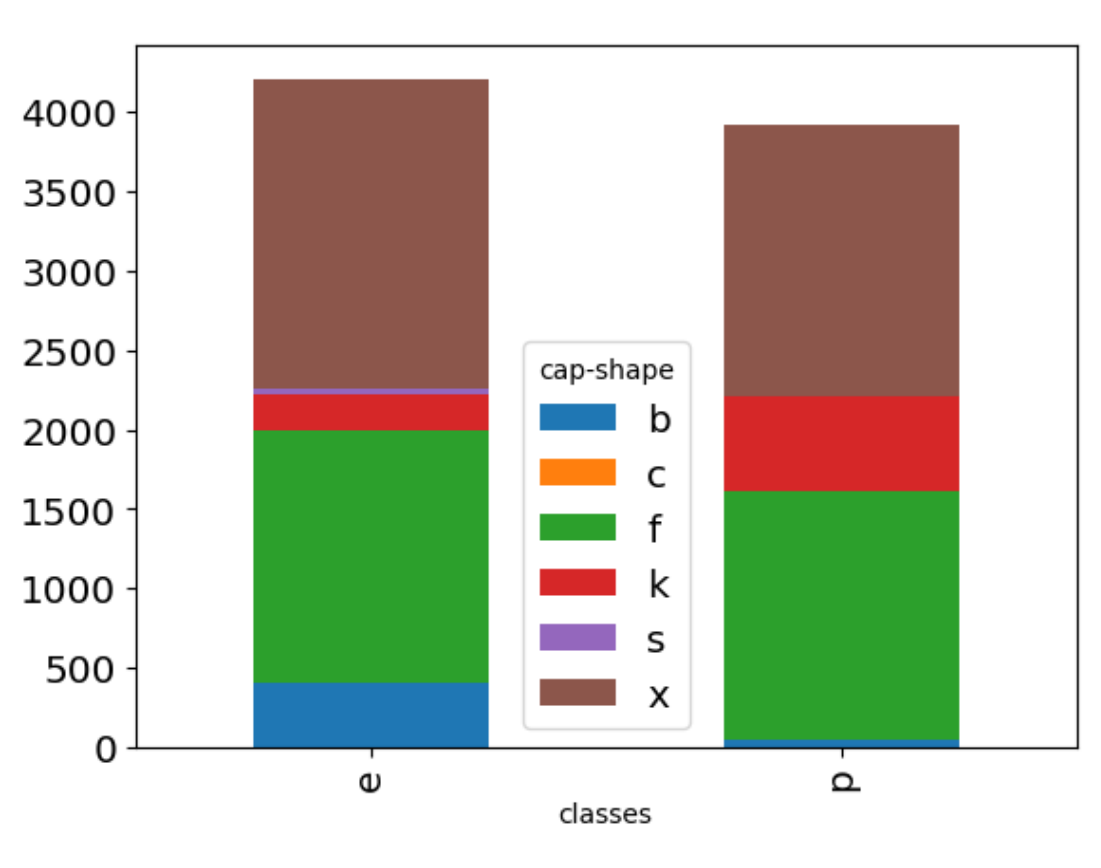
```
Chi-square value: 7659.726740165339
```

In [113…

```
pd.crosstab(df['classes'],df['cap-shape']).plot(kind='bar', stacked=True)
```

Out[113]:

```
<Axes: xlabel='classes'>
```

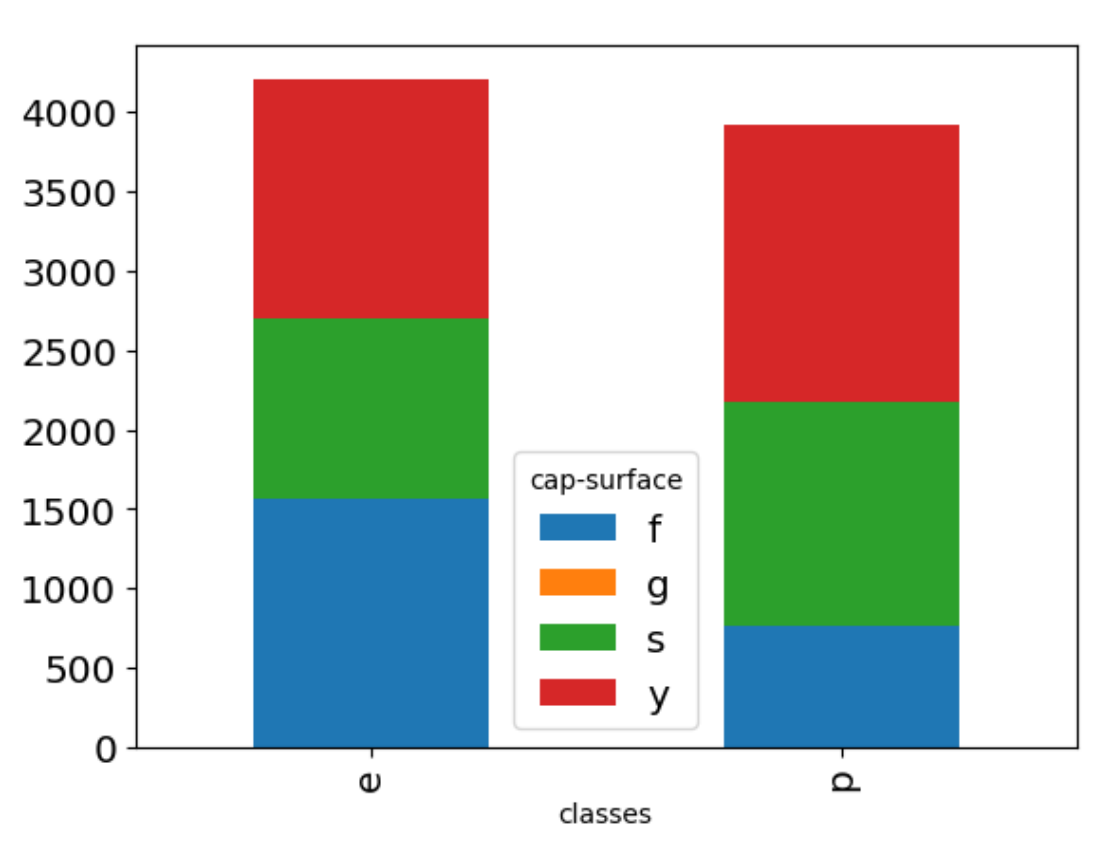


In [114…

```
pd.crosstab(df['classes'],df['cap-surface']).plot(kind='bar', stacked=True)
```

Out[114]:

```
<Axes: xlabel='classes'>
```

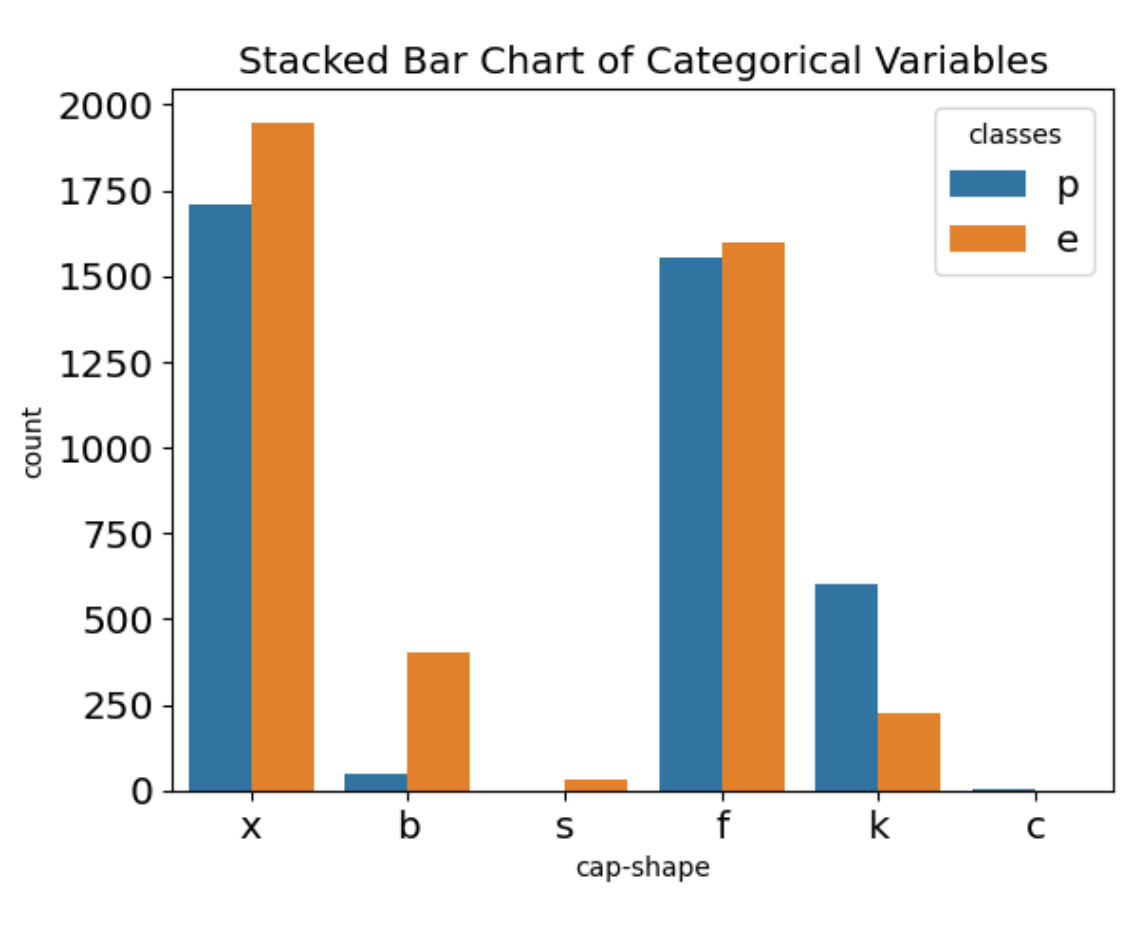


In [115…

```
sns.countplot(data=df, x='cap-shape', hue='classes')
plt.title('Stacked Bar Chart of Categorical Variables')
```

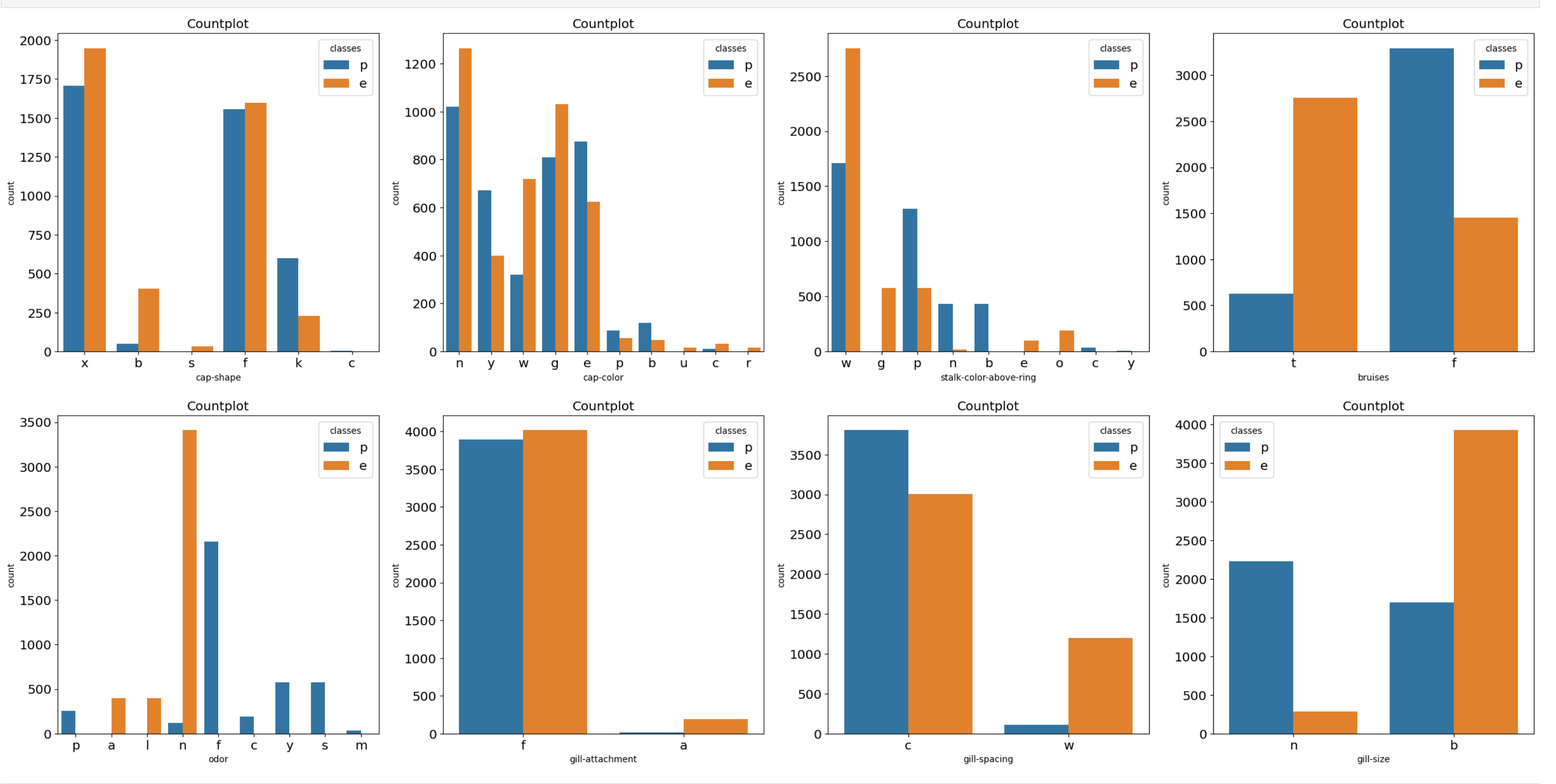
Out[115]:

```
Text(0.5, 1.0, 'Stacked Bar Chart of Categorical Variables')
```



In [135…

```
fig = plt.figure(figsize=(30,30))
axes1 = plt.subplot(4, 4, 1, title='Countplot')
sns.countplot(data=df, x='cap-shape', hue='classes')
axes2 = plt.subplot(4, 4, 2, title='Countplot')
sns.countplot(data=df, x='cap-color', hue='classes')
axes3 = plt.subplot(4, 4, 3, title='Countplot')
sns.countplot(data=df, x='stalk-color-above-ring', hue='classes')
axes4 = plt.subplot(4, 4, 4, title='Countplot')
sns.countplot(data=df, x='bruises', hue='classes')
axes5 = plt.subplot(4, 4, 5, title='Countplot')
sns.countplot(data=df, x='odor', hue='classes')
axes6 = plt.subplot(4, 4, 6, title='Countplot')
sns.countplot(data=df, x='gill-attachment', hue='classes')
axes7 = plt.subplot(4, 4, 7, title='Countplot')
sns.countplot(data=df, x='gill-spacing', hue='classes')
axes8 = plt.subplot(4, 4, 8, title='Countplot')
sns.countplot(data=df, x='gill-size', hue='classes')
plt.show()
```

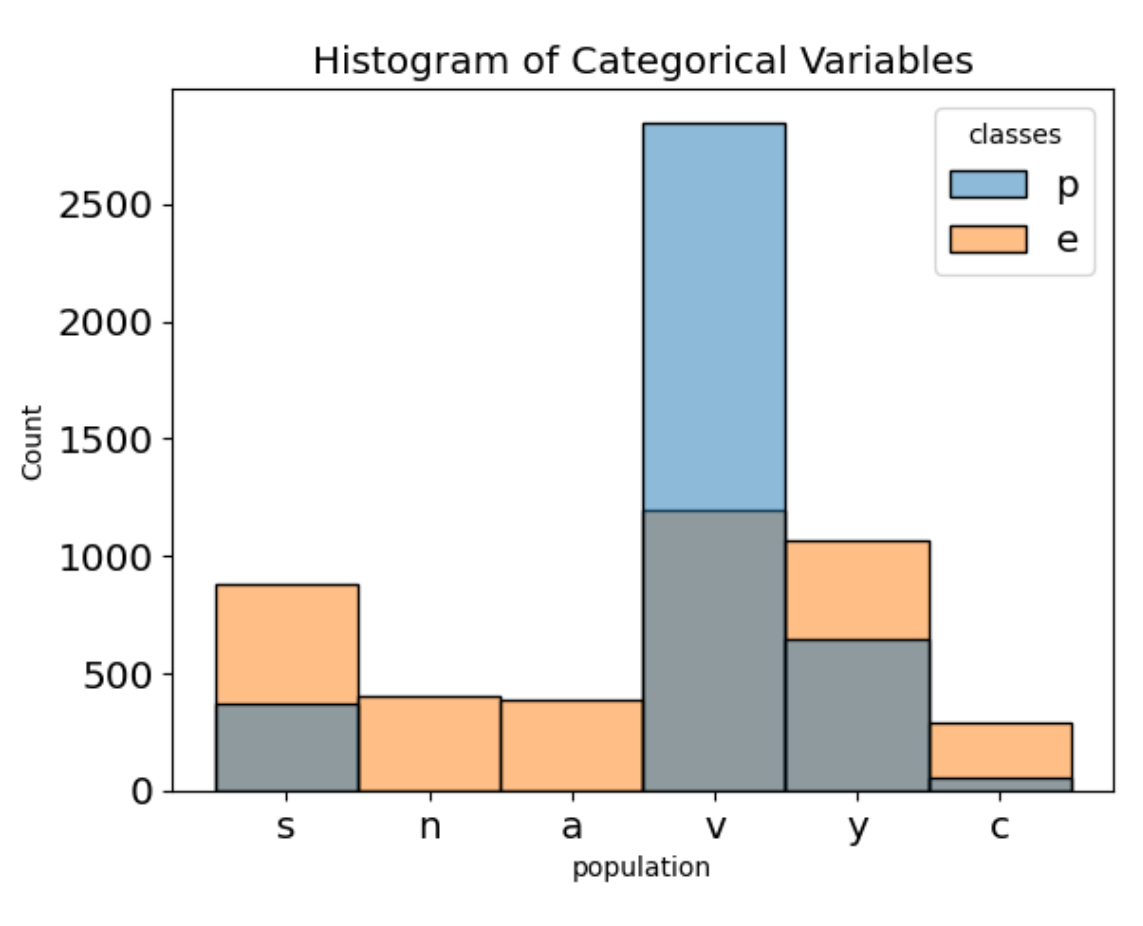


In [136…

```
sns.histplot(data=df, x="population", bins=30, hue="classes")
plt.title('Histogram of Categorical Variables')
```

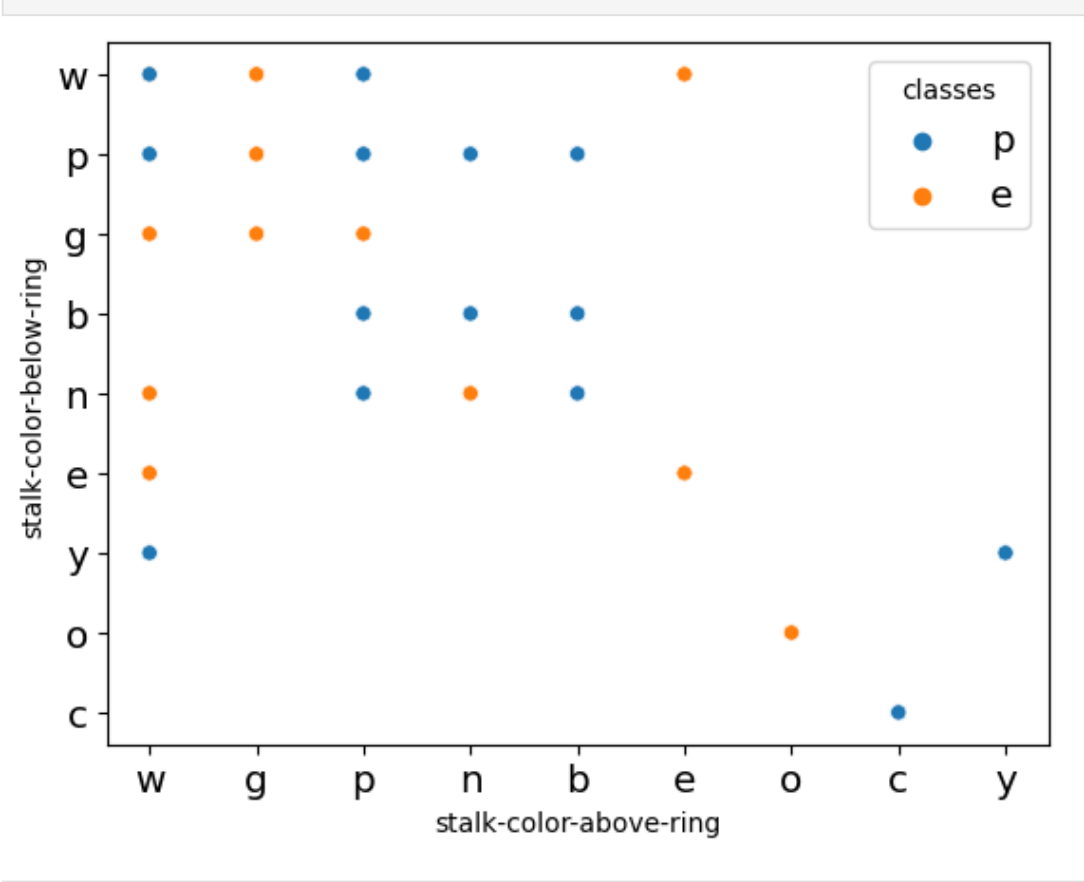
Out[136]:

```
Text(0.5, 1.0, 'Histogram of Categorical Variables')
```



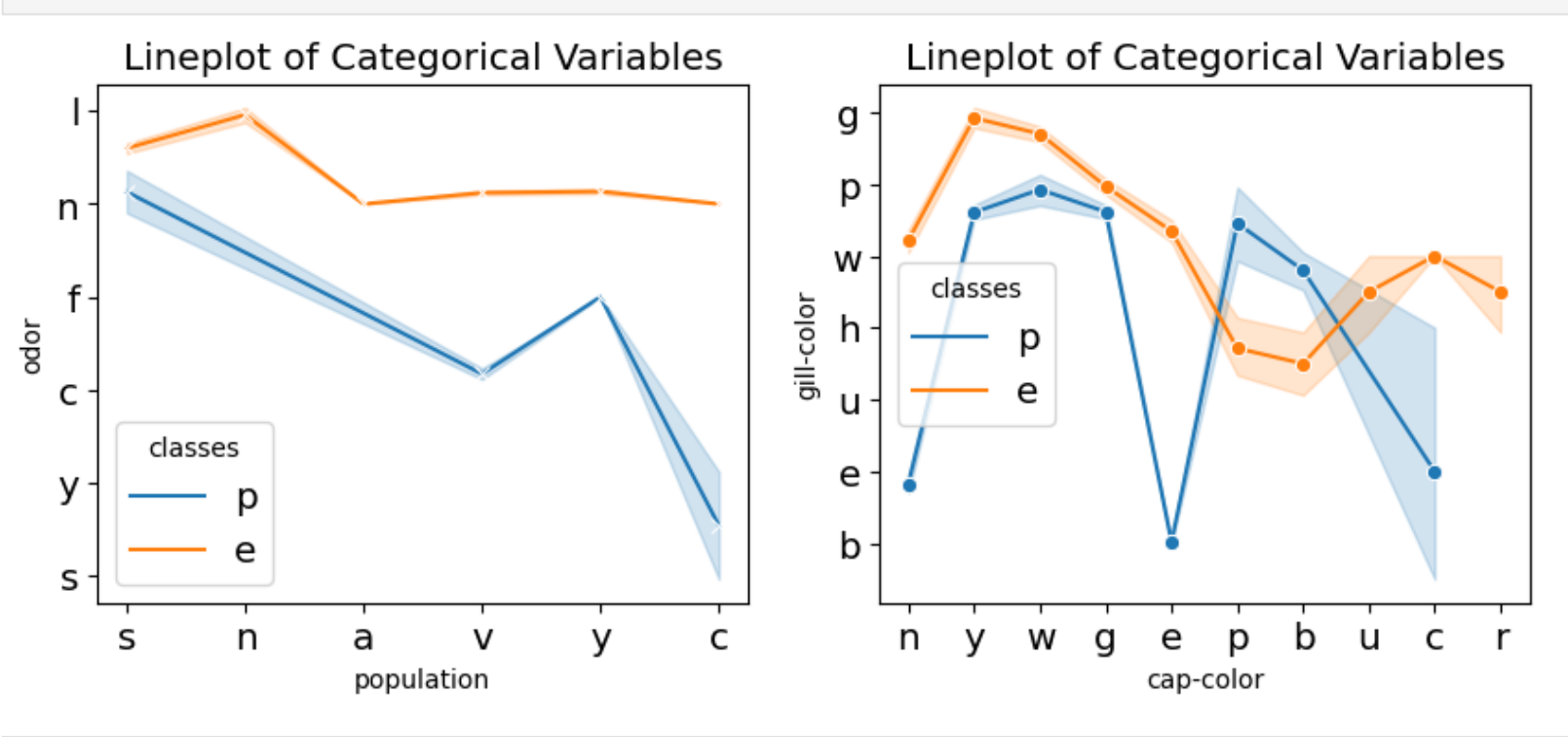
In [137…

```
sns.scatterplot(data=df, x="stalk-color-above-ring", y="stalk-color-below-ring", hue="classes")
plt.show()
```



In [138…

```
fig = plt.figure(figsize=(10,8))
axes1 = plt.subplot(2, 2, 1, title='Lineplot of Categorical Variables')
sns.lineplot(x="population", y ="odor", hue="classes", data=df, marker='x')
axes3 = plt.subplot(2, 2, 2, title='Plot2')
sns.lineplot(x="cap-color", y ="gill-color", hue="classes", data=df, marker='o')
plt.title('Lineplot of Categorical Variables')
plt.show()
```

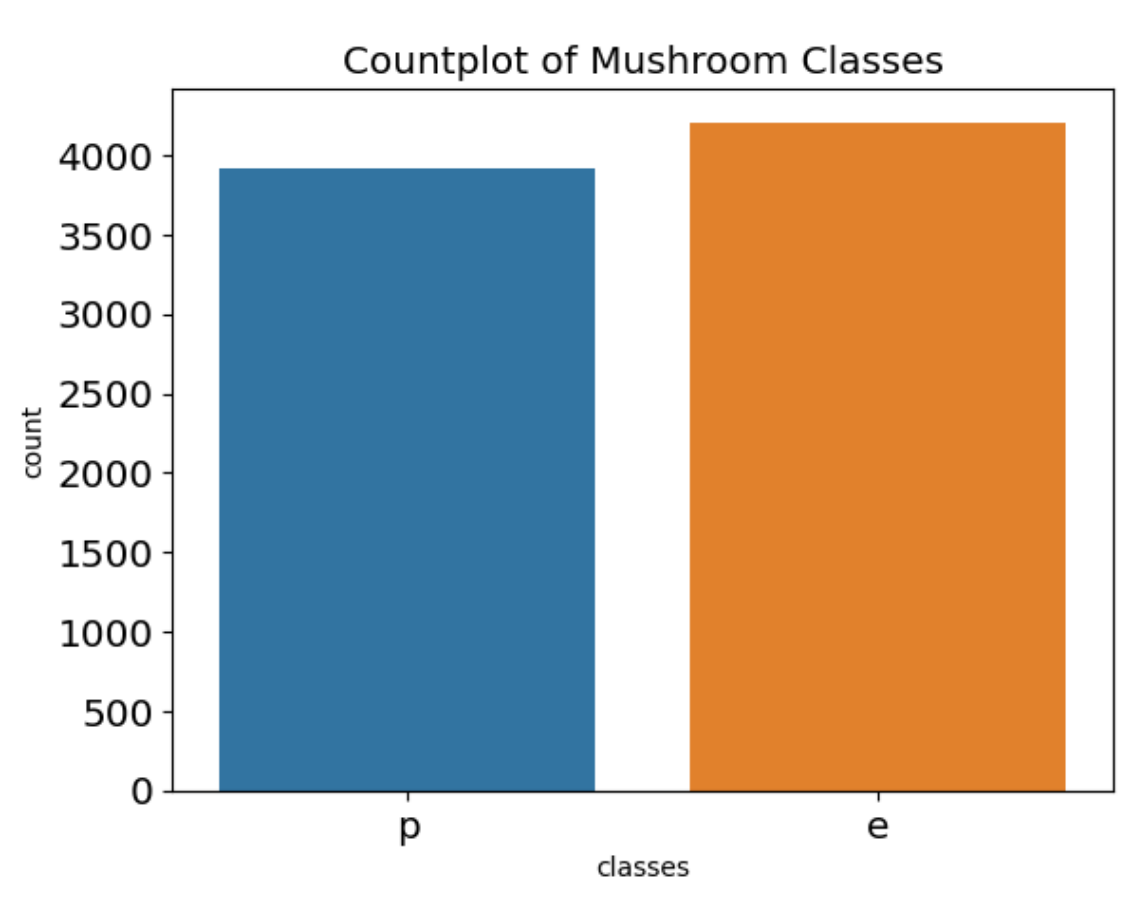


In [139…

```
sns.countplot(x ='classes', data=df);
plt.title('Countplot of Mushroom Classes')
```

Out[139]:

```
Text(0.5, 1.0, 'Countplot of Mushroom Classes')
```

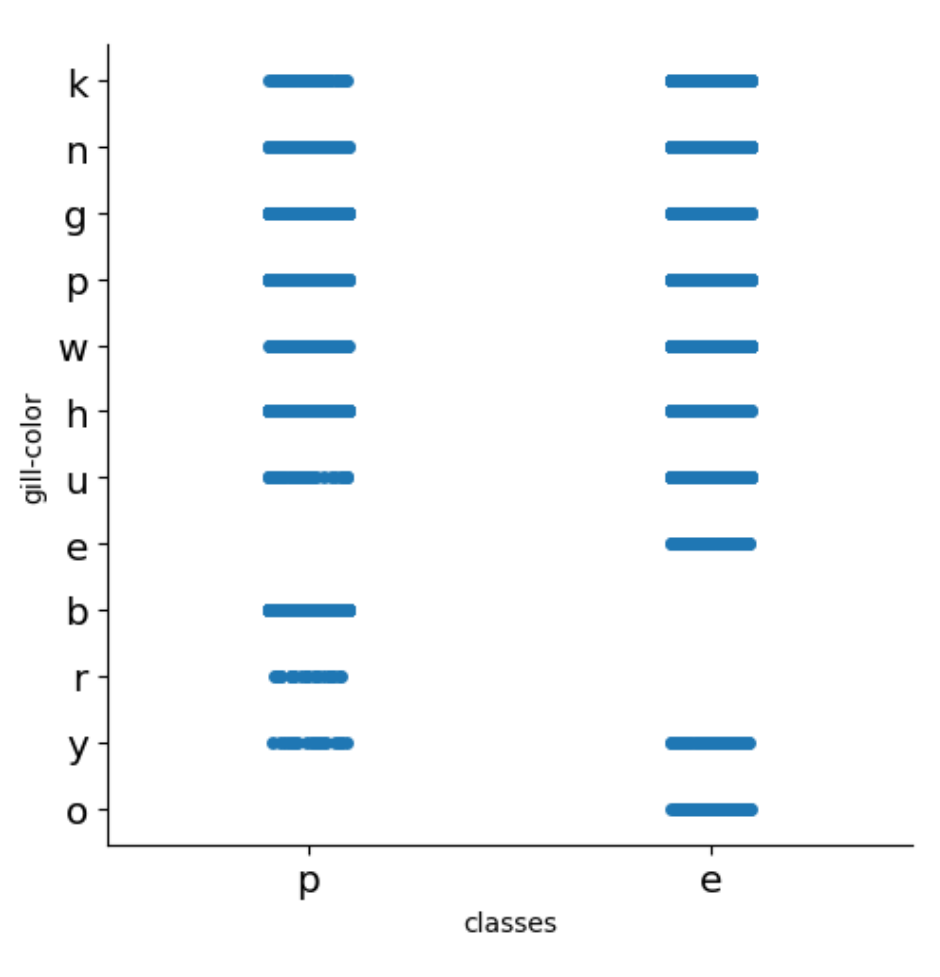


In [140…

```
sns.catplot(x='classes', y='gill-color', data=df)
```

Out[140]:

```
<seaborn.axisgrid.FacetGrid at 0x29d90851090>
```

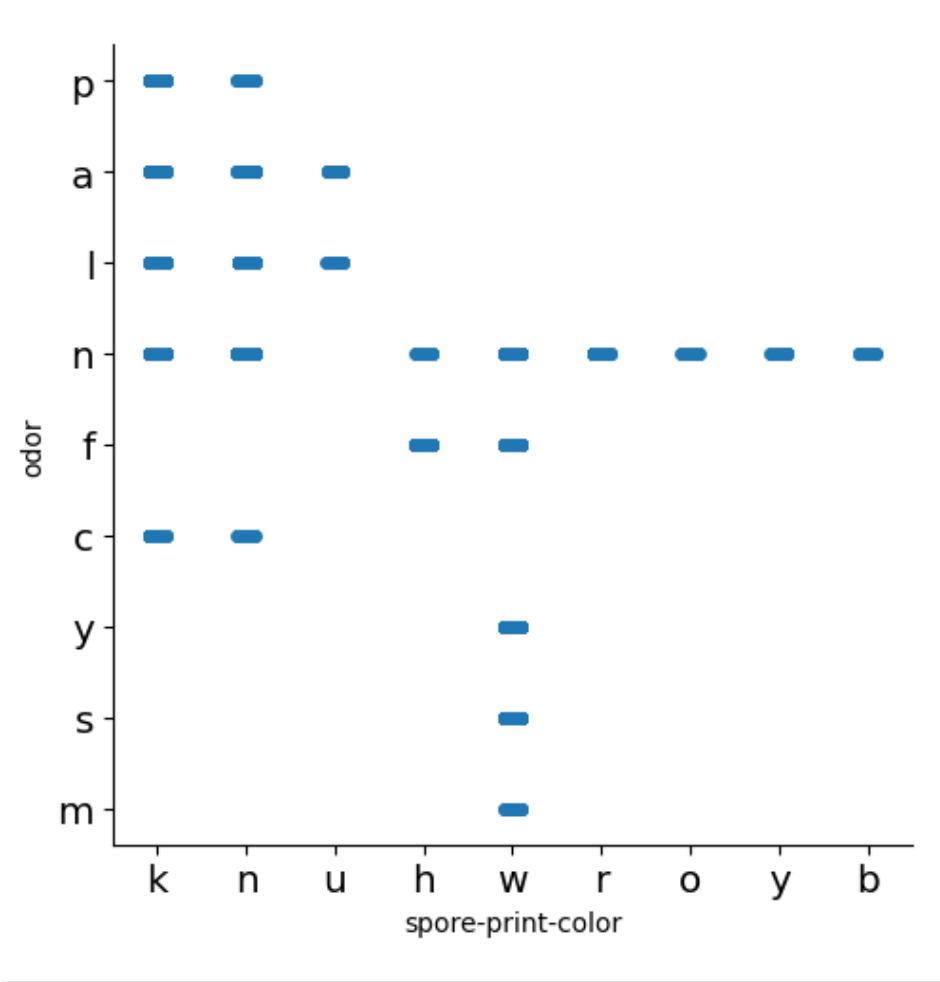


In [141…

```
sns.catplot(x='spore-print-color', y='odor', data=df)
```

Out[141]:

```
<seaborn.axisgrid.FacetGrid at 0x29d908209a0>
```

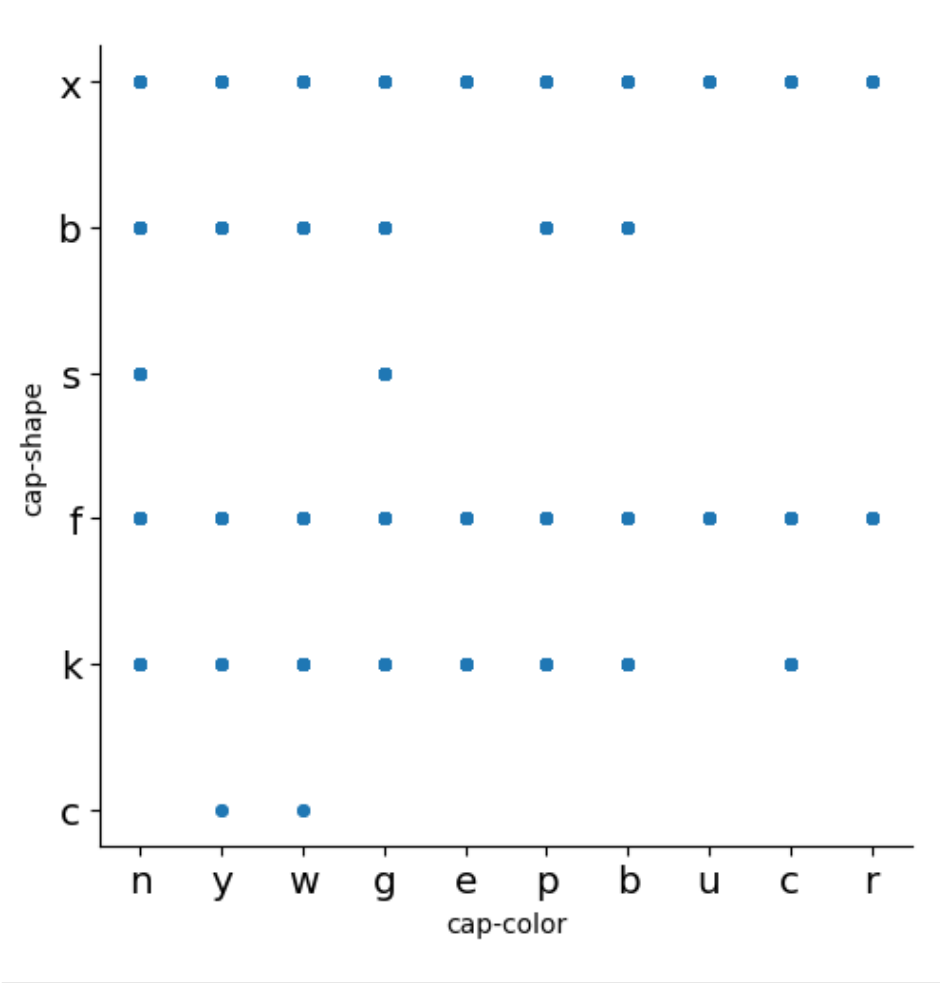


In [142…

```
sns.catplot(x='cap-color', y='cap-shape', jitter=False, data=df)
```

Out[142]:

```
<seaborn.axisgrid.FacetGrid at 0x29d90794790>
```

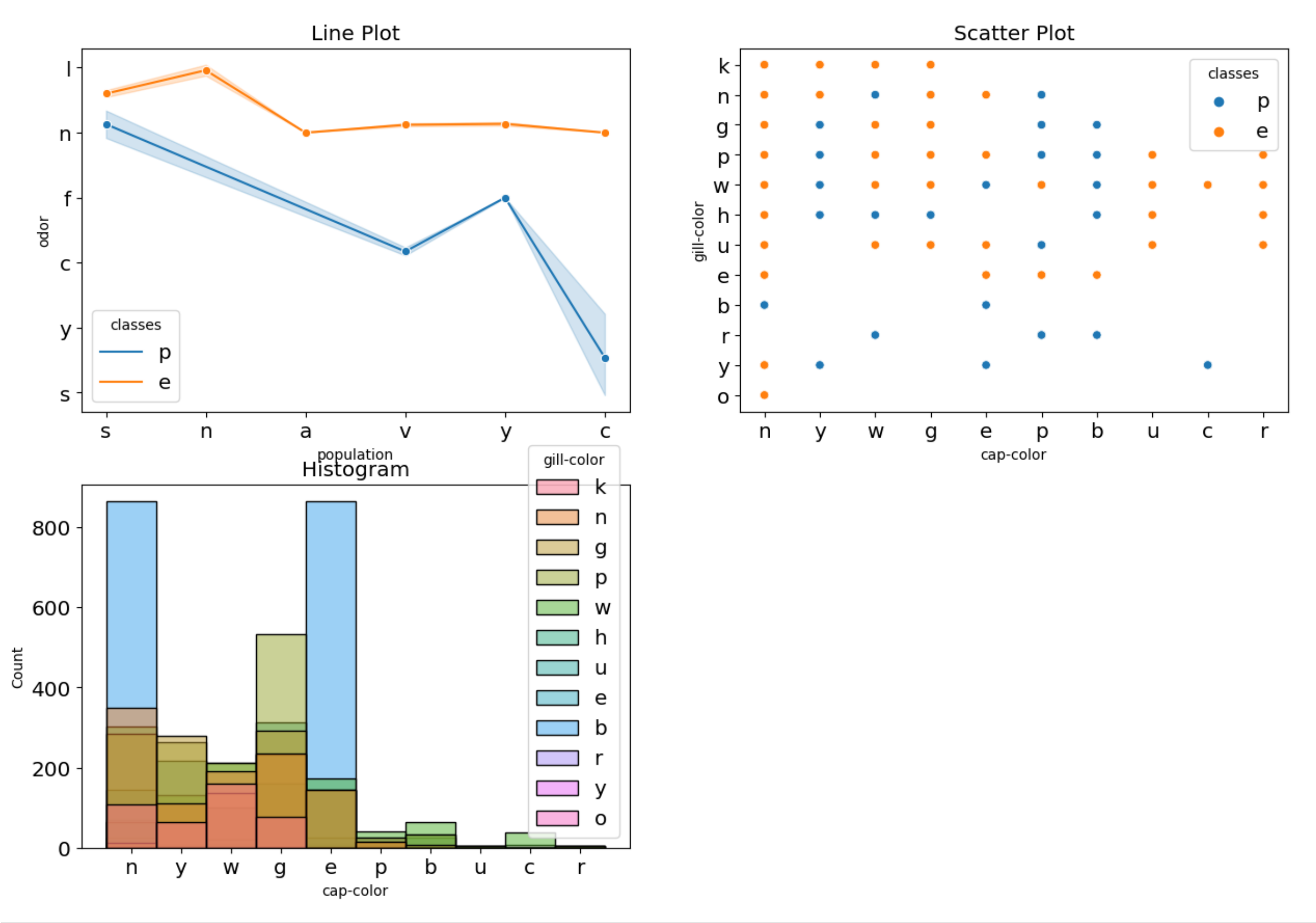


In [143…

```
fig = plt.figure(figsize=(15,10))
axes1 = plt.subplot(2, 2, 1, title='Line Plot')
sns.lineplot(x="population", y ="odor", hue="classes", data=df, marker='o')
axes2 = plt.subplot(2, 2, 2, title='Scatter Plot')
sns.scatterplot(data=df, x="cap-color", y="gill-color", hue="classes")
axes3 = plt.subplot(2, 2,3,title='Histogram')
sns.histplot(data=df, x="cap-color", bins=30, hue="gill-color")
```

Out[143]:

```
<Axes: title={'center': 'Histogram'}, xlabel='cap-color', ylabel='Count'>
```



In [144…

```
data = df.copy()
le = LabelEncoder()
for col in data.columns:
    data[col] = le.fit_transform(data[col])

data
```

Out[144]:

	classes	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	population	habitat
0	1	5	2	4	1	6	1	0	1	4	...	2	7	7	0	2	1	4	2	3	5
1	0	5	2	9	1	0	1	0	0	4	...	2	7	7	0	2	1	4	3	2	1
2	0	0	2	8	1	3	1	0	0	5	...	2	7	7	0	2	1	4	3	2	3
3	1	5	3	8	1	6	1	0	1	5	...	2	7	7	0	2	1	4	2	3	5
4	0	5	2	3	0	5	1	1	0	4	...	2	7	7	0	2	1	0	3	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8119	0	3	2	4	0	5	0	0	0	11	...	2	5	5	0	1	1	4	0	1	2
8120	0	5	2	4	0	5	0	0	0	11	...	2	5	5	0	0	1	4	0	4	2
8121	0	2	2	4	0	5	0	0	0	5	...	2	5	5	0	1	1	4	0	1	2
8122	1	3	3	4	0	8	1	0	1	0	...	1	7	7	0	2	1	0	7	4	2
8123	0	5	2	4	0	5	0	0	0	11	...	2	5	5	0	1	1	4	4	1	2

8124 rows × 23 columns

In [145…

```
data.describe()
```

Out[145]:

	classes	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	population	habitat
count	8124.000000	8124.000000	8124.000000	8124.000000	8124.000000	8124.000000	8124.000000	8124.000000	8124.000000	8124.000000	...	8124.000000	8124.000000	8124.000000	8124.0	8124.000000	8124.000000	8124.000000	8124.000000	8124.000000	8124.000000
mean	0.482029	3.348104	1.827671	4.504677	0.415559	4.144756	0.974151	0.161497	0.309207	4.810684	...	1.603644	5.816347	5.794682	0.0	1.965534	1.069424	2.291974	3.596750	3.644018	1.508616
std	0.499708	1.604329	1.229873	2.545821	0.492848	2.103729	0.158695	0.368011	0.462195	3.540359	...	0.675974	1.901747	1.907291	0.0	0.242669	0.271064	1.801672	2.382663	1.252082	1.719975
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	0.000000	3.000000	0.000000	2.000000	1.000000	0.000000	0.000000	2.000000	...	1.000000	6.000000	6.000000	0.0	2.000000	1.000000	0.000000	2.000000	3.000000	0.000000
50%	0.000000	3.000000	2.000000	4.000000	0.000000	5.000000	1.000000	0.000000	0.000000	5.000000	...	2.000000	7.000000	7.000000	0.0	2.000000	1.000000	2.000000	3.000000	4.000000	1.000000
75%	1.000000	5.000000	3.000000	8.000000	1.000000	5.000000	1.000000	0.000000	1.000000	7.000000	...	2.000000	7.000000	7.000000	0.0	2.000000	1.000000	4.000000	7.000000	4.000000	2.000000
max	1.000000	5.000000	3.000000	9.000000	1.000000	8.000000	1.000000	1.000000	1.000000	11.000000	...	3.000000	8.000000	8.000000	0.0	3.000000	2.000000	4.000000	8.000000	5.000000	6.000000

8 rows × 23 columns

In [146…

```
sns.boxplot(data = data, orient='h')
```

Out[146]:

```
<Axes: >
```