Week 3 Quiz

Qi Meng - qm2162

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

sns.set_style('darkgrid')

%matplotlib inline
```

In [2]:
```
# Use pandas to read in 'wine_dataset.csv'
# This is a dataset of various wines with a target of categorical variable 'c
df = pd.read_csv('../data/wine_dataset.csv')
```

In [3]:
```
# 1. Print out the number of rows and columns in the dataset using shape
df.shape
```

Out[3]:
```
(178, 14)
```

In [4]:
```
# 2. Display the first 3 rows of df using head
df.head(3)
```

Out[4]:

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavan
0	14.23	1.71	2.43	15.6	127.0	2.80	3.06	
1	13.20	1.78	2.14	11.2	100.0	2.65	2.76	
2	13.16	2.36	2.67	18.6	101.0	2.80	3.24	

In [5]:
```
# 3. Display the summary stats of numeric columns using df.describe()
df.describe()
```

Out[5]:

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flava
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.00
mean	13.000618	2.336348	2.366517	19.494944	99.741573	2.295112	2.02
std	0.811827	1.117146	0.274344	3.339564	14.282484	0.625851	0.99
min	11.030000	0.740000	1.360000	10.600000	70.000000	0.980000	0.34
25%	12.362500	1.602500	2.210000	17.200000	88.000000	1.742500	1.20
50%	13.050000	1.865000	2.360000	19.500000	98.000000	2.355000	2.13
75%	13.677500	3.082500	2.557500	21.500000	107.000000	2.800000	2.87
max	14.830000	5.800000	3.230000	30.000000	162.000000	3.880000	5.08

In [6]:

```
# 4. Using iloc, display the first 3 rows, first 3 columns
# You should see the columns ['alcohol','malic_acid','ash']

df.iloc[:3,:3]
```

Out[6]:

	alcohol	malic_acid	ash
0	14.23	1.71	2.43
1	13.20	1.78	2.14
2	13.16	2.36	2.67

In [7]:

```
# 5. Using loc, display rows with index 4 to 6 inclusive and columns 'ash' an
df.loc[4:6,['ash','total_phenols']]
```

Out[7]:

	ash	total_phenols
4	2.87	2.80
5	2.45	3.27
6	2.45	2.50

In [8]:

```
# 6. Display 'ash' and 'hue' columns for all rows with 'hue' greater than the
# This should result in 89 rows x 2 columns
df.loc[df['hue']> df['hue'].median(), ['ash', 'hue']]
```

Out[8]:

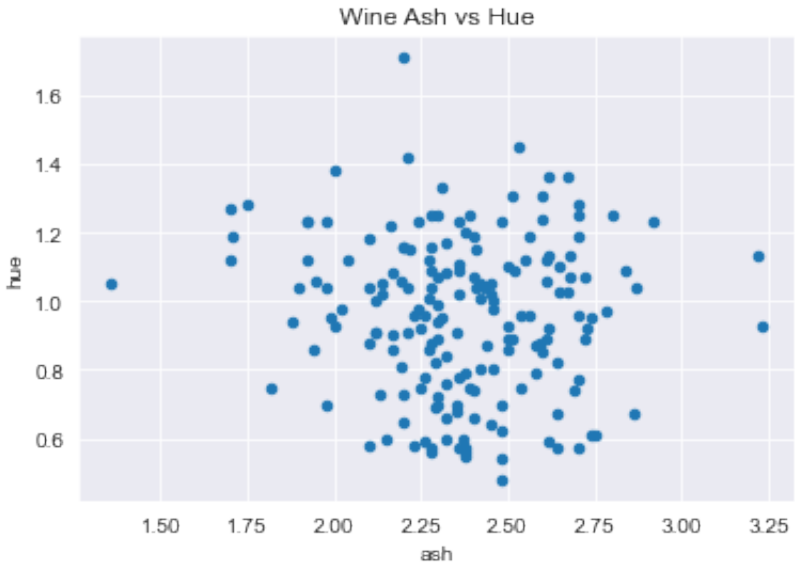
	ash	hue
0	2.43	1.04
1	2.14	1.05
2	2.67	1.03
4	2.87	1.04
5	2.45	1.05
...	...	...
112	2.92	1.23
113	2.50	1.10
115	2.20	1.71
117	2.19	1.06
127	2.78	0.97

89 rows × 2 columns

In [9]:

```
# 7.1 Create a scatterplot of 'ash' and 'hue' for all rows in df using df.plo
# Place ash on the x-axis and hue on the y-axis.
# Store the returned output in ax.
ax = df.plot.scatter(x='ash', y='hue');

# 7.2 Add the title 'Wine Ash vs Hue' to the plot using ax.set_title()
ax.set_title('Wine Ash vs Hue');
```



In [15]:

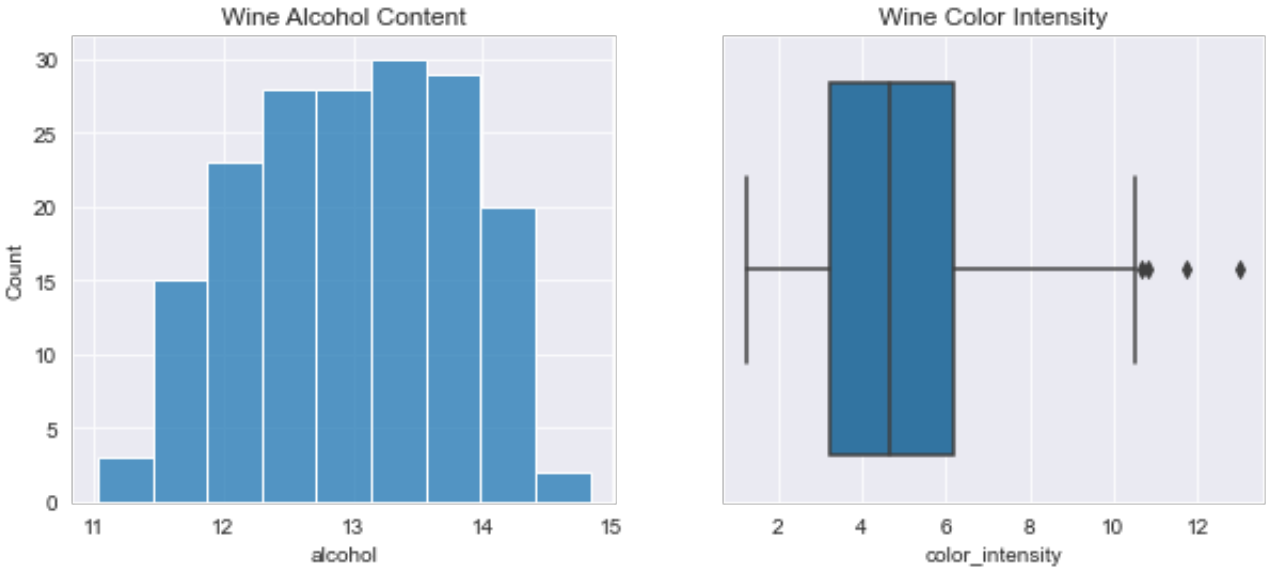
```
# 8.1. Create two axes using plt.subplots with 1 row , 2 columns, and figsize
fig,ax = plt.subplots(1, 2, figsize=(10,4));

# 8.2 In the first axis (ax[0]), plot the distribution of df.alcohol using sn
sns.histplot(x='alcohol', data=df, ax=ax[0]);

# 8.3 Add the title 'Wine Alcohol Content' to ax[0] using set_title()
ax[0].set_title('Wine Alcohol Content');

# 8.4 In the second axis (ax[1]), plot a boxplot of df.color_intensity using
sns.boxplot(x='color_intensity', data=df, ax=ax[1]);

# 8.5 Add the title 'Wine Color Intensity' to ax[1] using set_title()
ax[1].set_title('Wine Color Intensity');
```



In [11]:

```
# when completed,
#   make sure you've replaced Name and UNI in the first cell and filename
#   use Print Preview, Print-> Save to pdf
#   and post pdf to GradeScope
```