Chem 277B Spring 2024 Tutorial 5

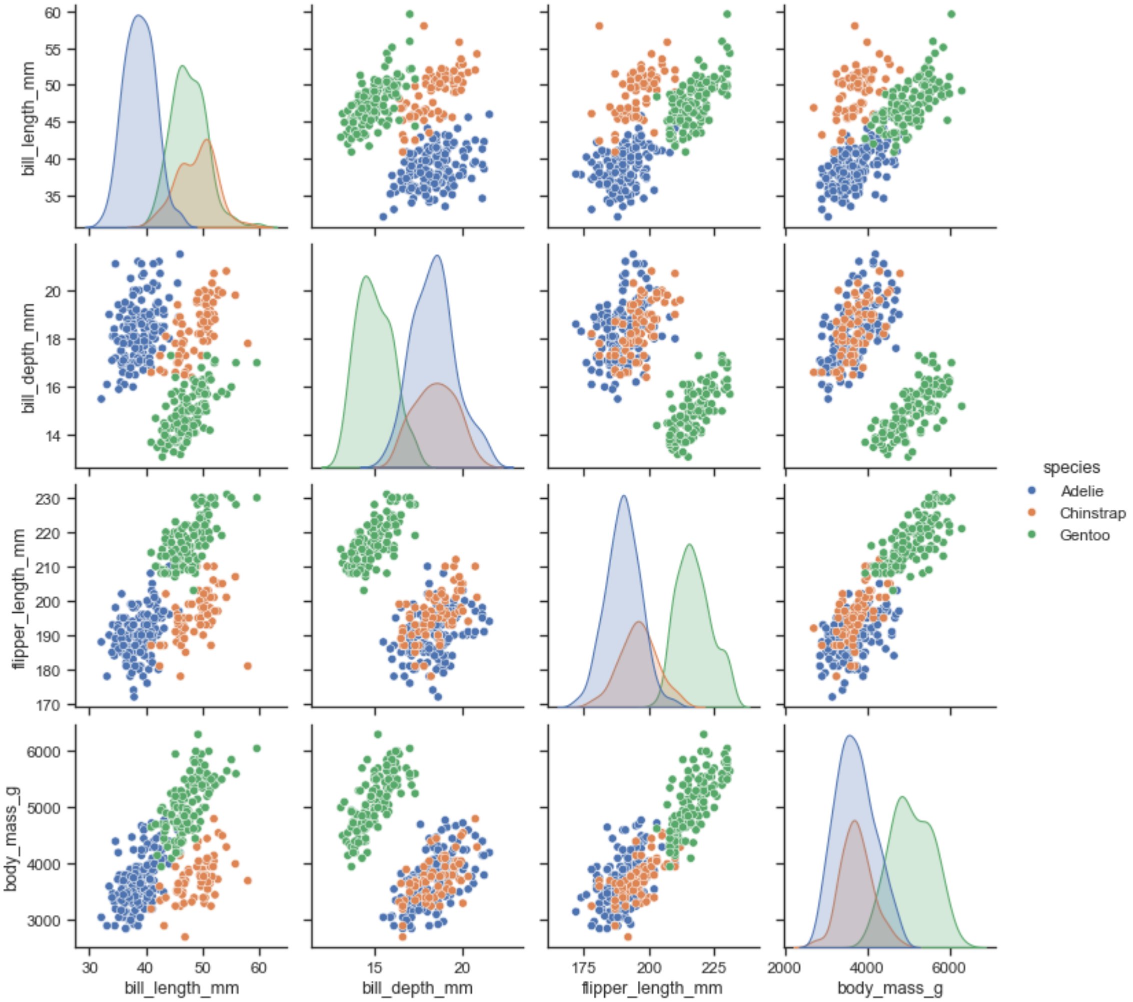
Outline

- Introduction and installation
- Data visualization with `catplot`, `relplot` and `pairplot`
- Correlation matrix
- Clustering
- Quick Markdown & LaTeX syntax

Seaborn

- Documentation
- Installation: `pip install seaborn` or `conda install seaborn -c conda-forge`

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn also provides better integration with Pandas data structures.



Bar Plot

In [1]:

```
import pandas as pd
import seaborn as sns
```

```
/var/folders/k8/mg372j_55z30k1z4y_8mb0w00000gn/T/ipykernel_34432/432526209.p
y:1: DeprecationWarning:
Pyarrow will become a required dependency of pandas in the next major releas
e of pandas (pandas 3.0),
(to allow more performant data types, such as the Arrow string type, and bet
ter interoperability with other libraries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at https://github.com/pandas-dev/pandas/issues/54
466

  import pandas as pd
```

In [3]:

```
df = pd.read_csv("compounds.csv")
df.head()
```

Out[3]:

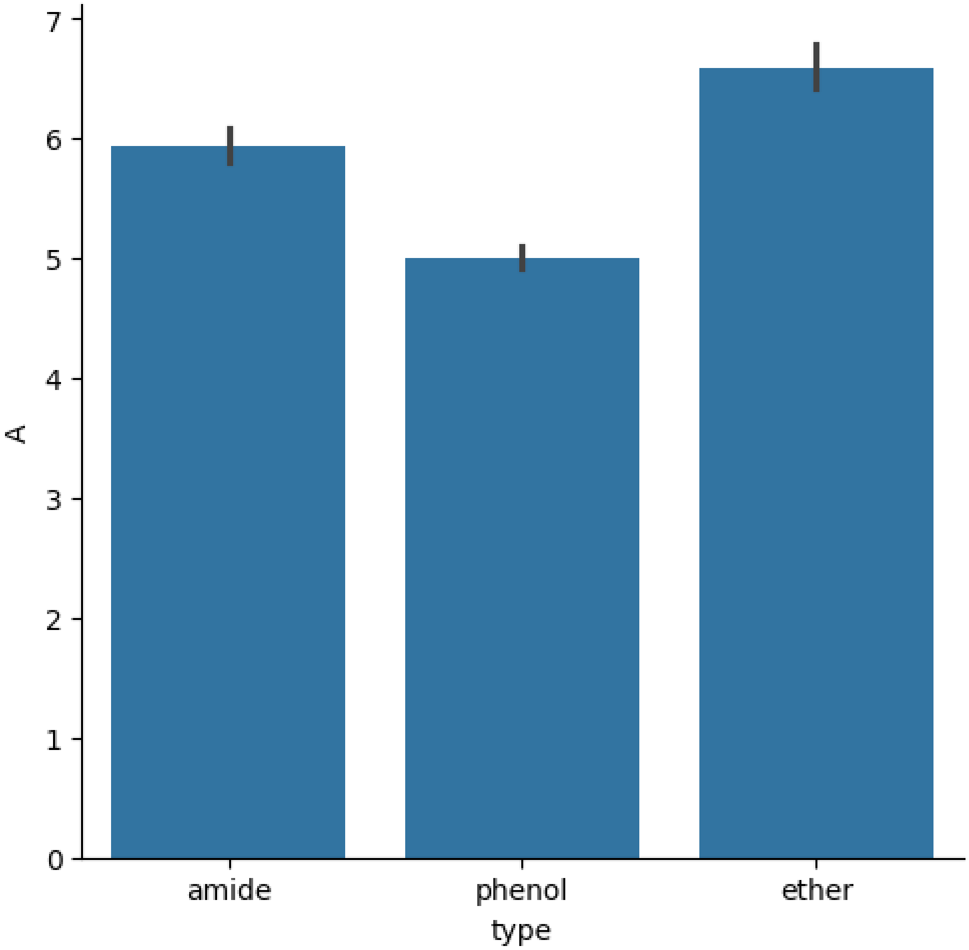
	A	B	C	D	type	Start assignment
0	6.4	2.9	4.3	1.3	amide	1
1	5.7	4.4	1.5	0.4	phenol	2
2	6.7	3.0	5.2	2.3	ether	0
3	5.8	2.8	5.1	2.4	ether	1
4	6.4	3.2	5.3	2.3	ether	0

In [11]:

```
sns.catplot(data=df, x='type', y='A', kind='bar')
```

Out[11]:

```
<seaborn.axisgrid.FacetGrid at 0x29ab24490>
```



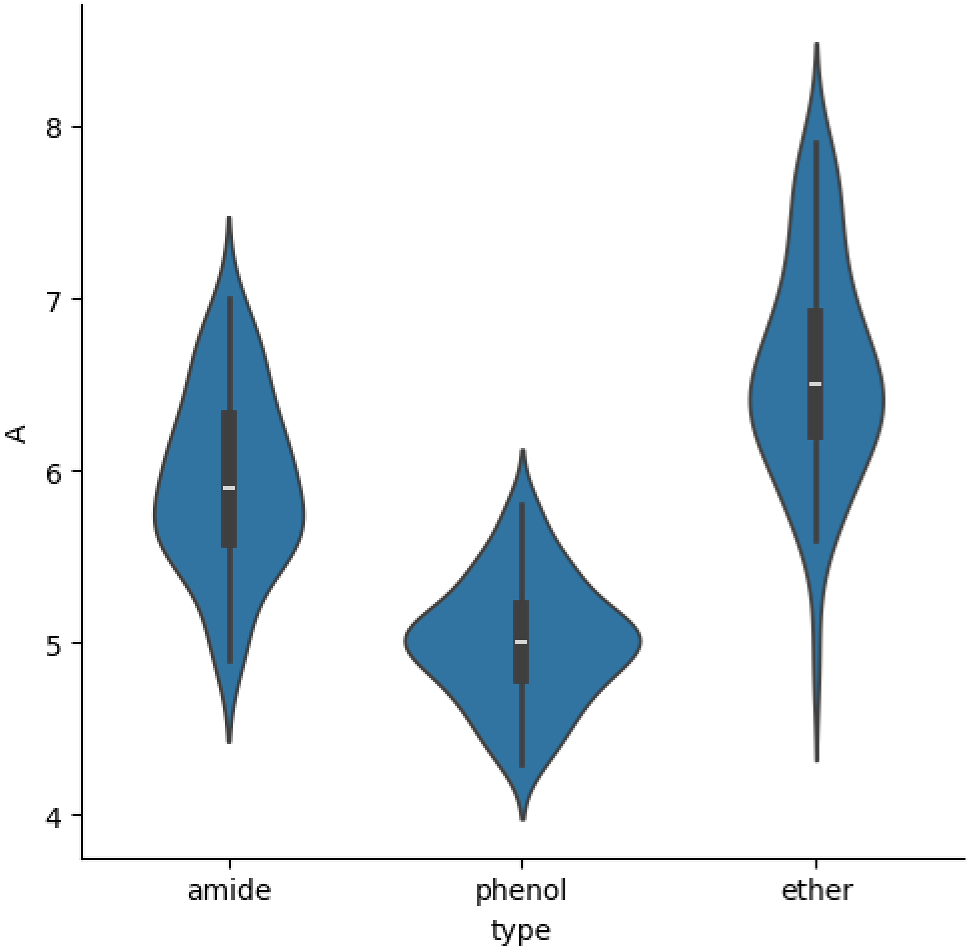
Violin plot

In [12]:

```
sns.catplot(data=df, x='type', y='A', kind='violin')
```

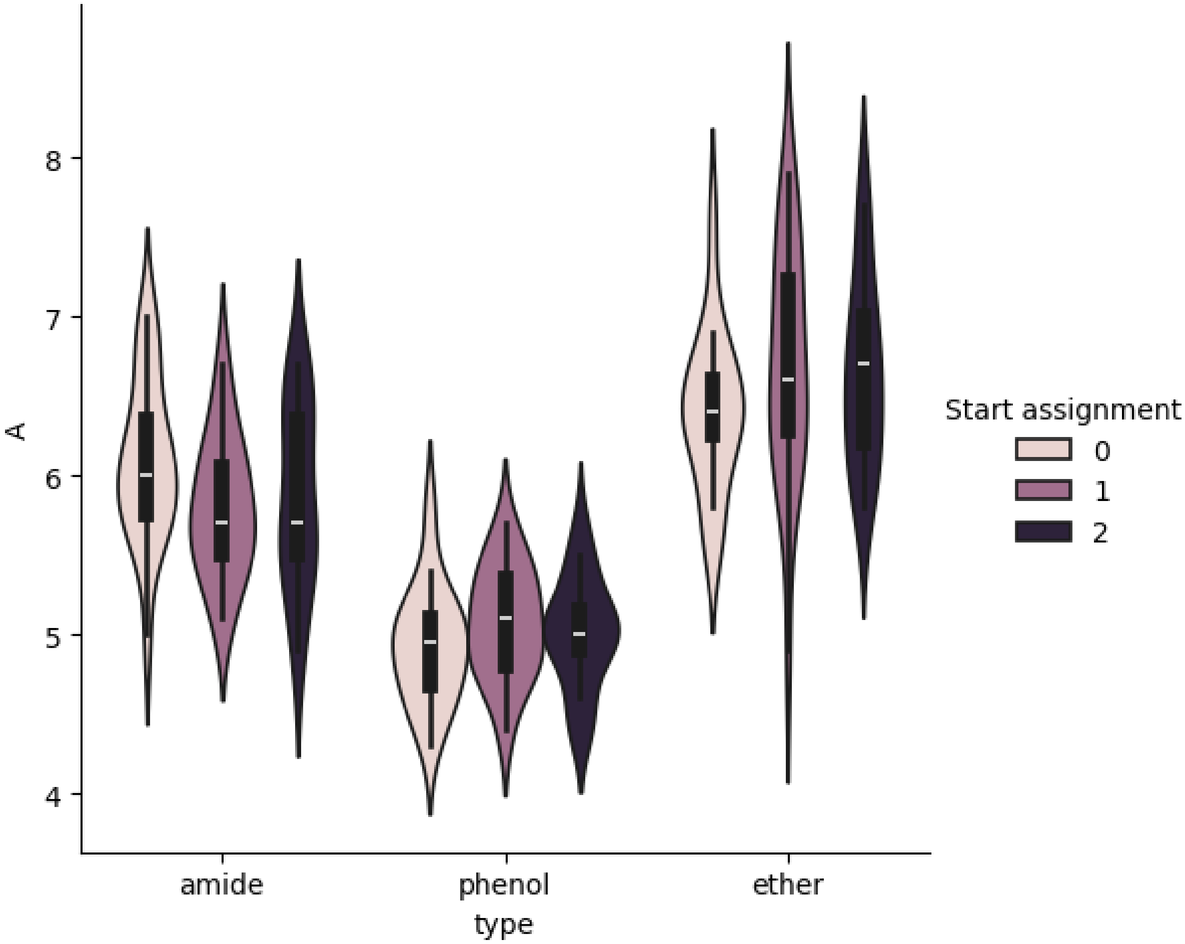
```
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
```

```
Out[12]: <seaborn.axisgrid.FacetGrid at 0x29abd3be0>
```



```
In [6]: sns.catplot(data=df, x='type', y='A', kind='violin', hue='Start assignment')
```

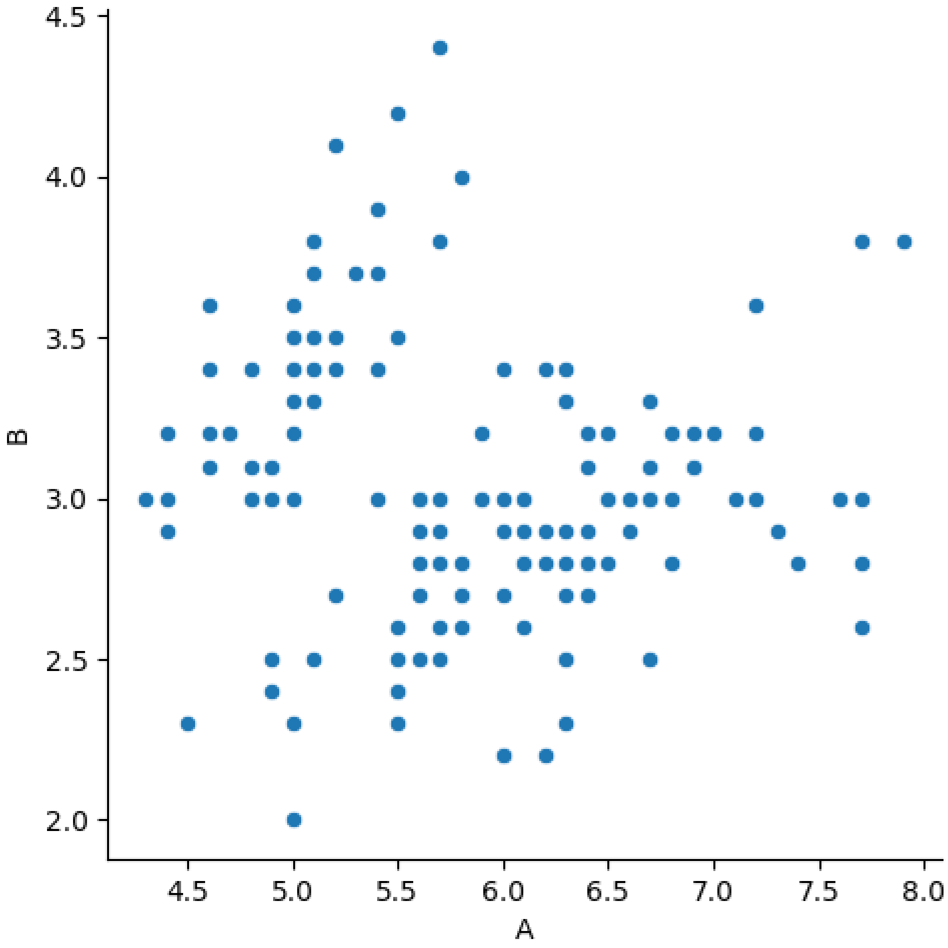
```
Out[6]: <seaborn.axisgrid.FacetGrid at 0x2977195a0>
```



Scatter plot

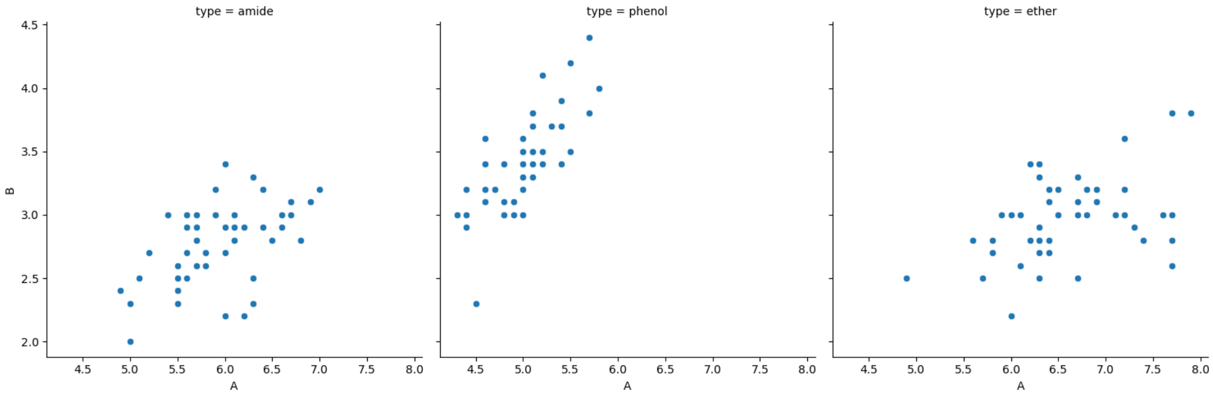
```
In [13]: sns.relplot(data=df, x='A', y='B')
```

```
Out[13]: <seaborn.axisgrid.FacetGrid at 0x29ac57d90>
```



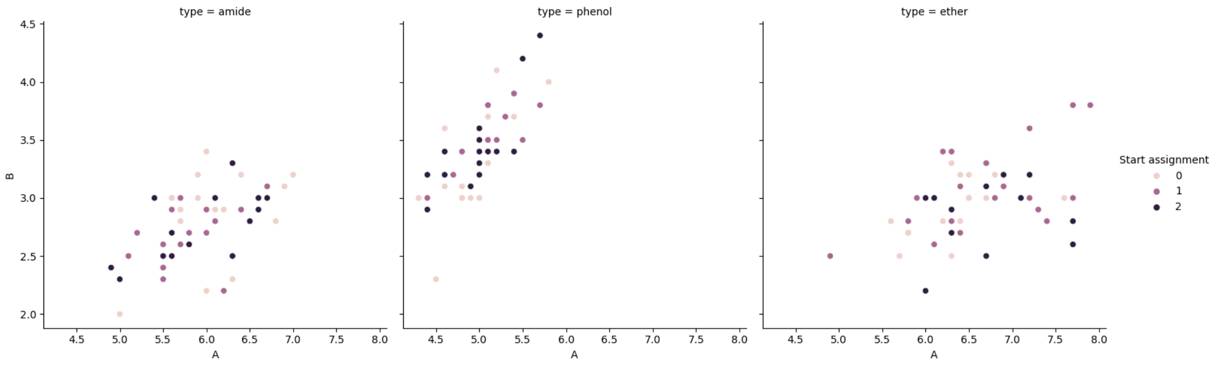
In [14]:
```
sns.relplot(data=df, x='A', y='B', col='type')
```

Out[14]:
```
<seaborn.axisgrid.FacetGrid at 0x29ad14910>
```



In [15]:
```
sns.relplot(data=df, x='A', y='B', col='type', hue='Start assignment')
```

Out[15]:
```
<seaborn.axisgrid.FacetGrid at 0x29ae7a2c0>
```



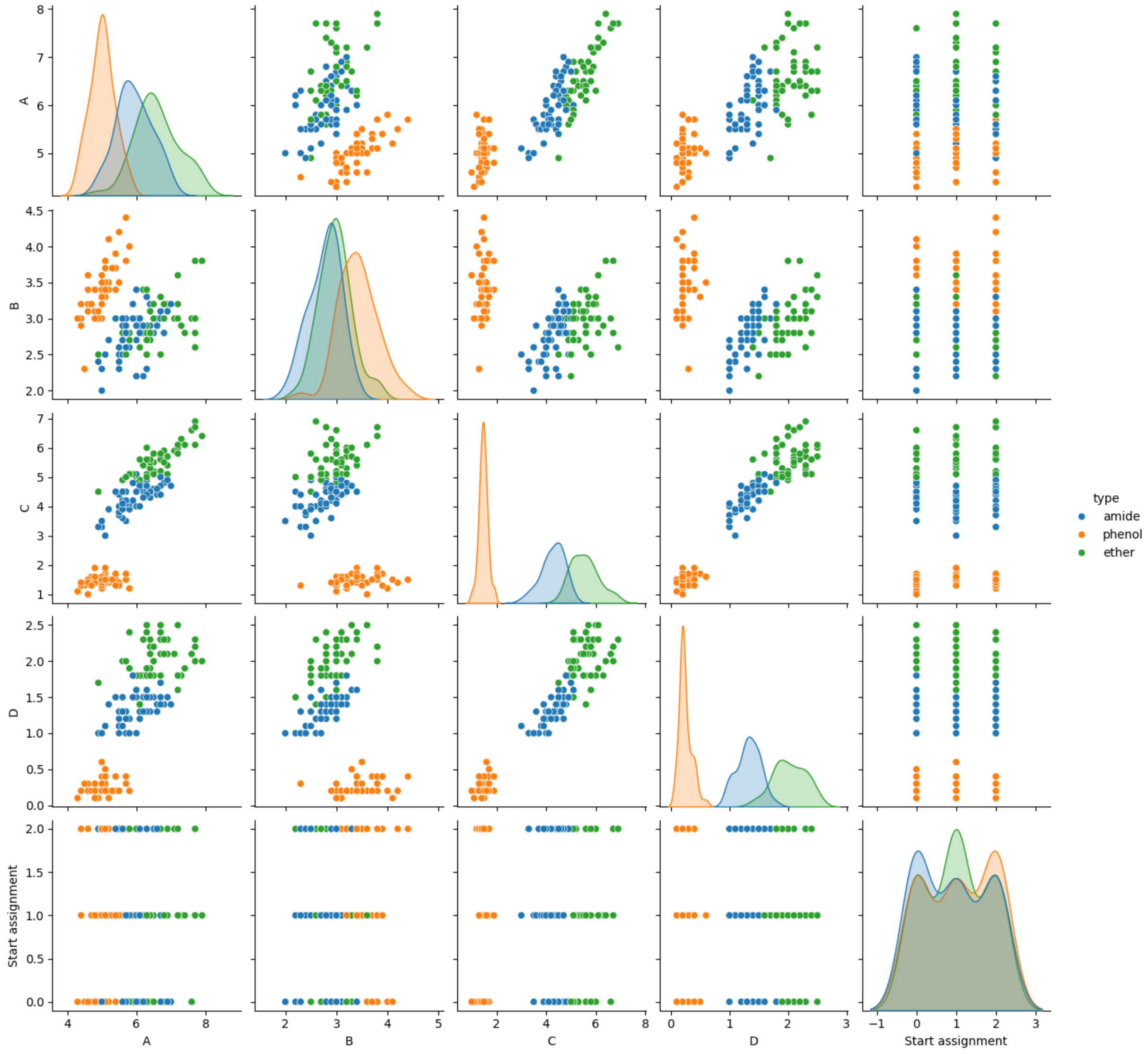
Pair Plot

In [17]:
```
sns.pairplot(data=df, hue='type')
```

```
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
```

```
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
/Users/chu/miniconda3/envs/chem277b/lib/python3.10/site-packages/seaborn/_ba
se.py:949: FutureWarning: When grouping with a length-1 list-like, you will
need to pass a length-1 tuple to get_group in a future version of pandas. Pa
ss `(name,)` instead of `name` to silence this warning.
  data_subset = grouped_data.get_group(pd_key)
```

```
Out[17]:  <seaborn.axisgrid.PairGrid at 0x29baad960>
```



Correlation matrix

The **correlation coefficient** between two random variables X and Y are defined as:

$$\frac{\mathrm{Cov}(\mathrm{X},\mathrm{Y})}{\sqrt{\mathrm{Var}(\mathrm{X})\mathrm{Var}(\mathrm{Y})}} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{[\sum_i (X_i - \bar{X})^2][\sum_i (Y_i - \bar{Y})^2]}}$$

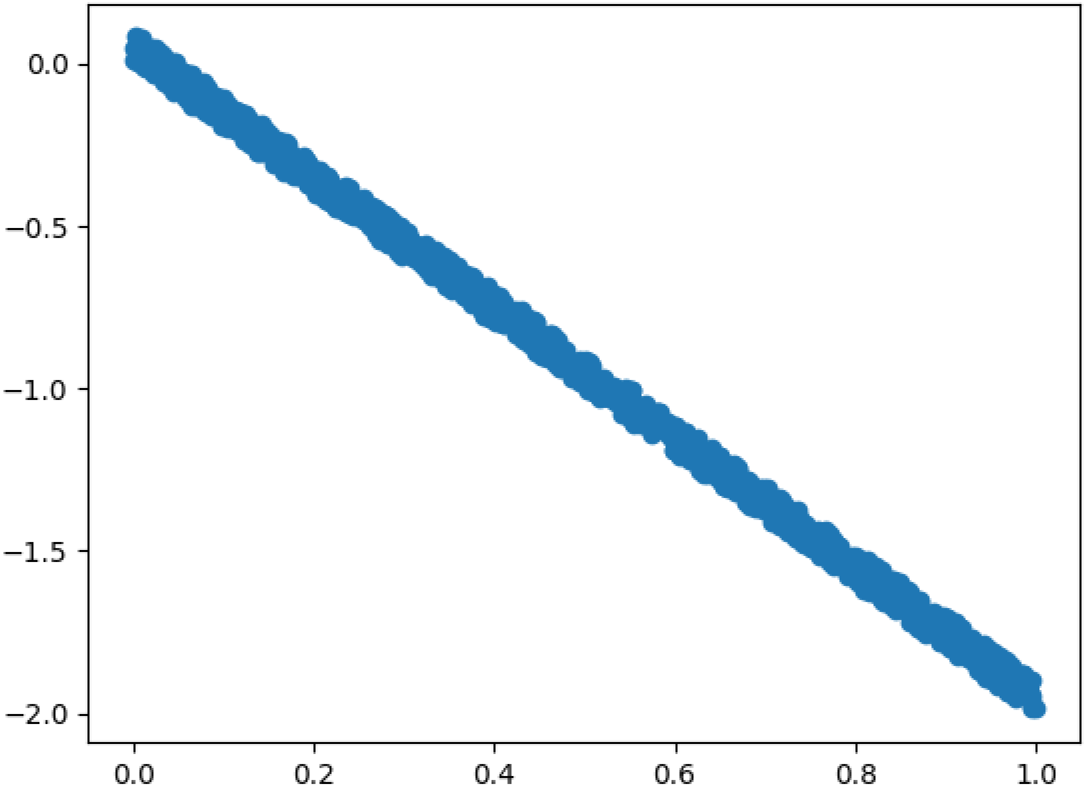
The correlation coefficient should be in the range of $[-1, 1]$. When $X = Y$, the correlation coefficient will be 1, and when $X = -Y$ the correlation coefficient will be -1.

In [24]:

```python
import numpy as np, matplotlib.pyplot as plt


X = np.random.random(1000)
# Y = np.random.random(1000)
Y = -2*X + np.random.random(1000)*.1

def corrcoef(X, Y):
    X_shift = X - np.mean(X)
    Y_shift = Y - np.mean(Y)
    return np.sum(X_shift * Y_shift) / np.sqrt(np.sum(X_shift ** 2) * np.sum

plt.scatter(X, Y)
corrcoef(X, Y)
```

Out[24]:

```
-0.9988434853741592
```



In [25]:
```
wines = pd.read_csv("wines.csv").iloc[:, :-2]
features = wines.values
features.shape
```

Out[25]:
```
(178, 13)
```

In [28]:
```
wines.head()
```

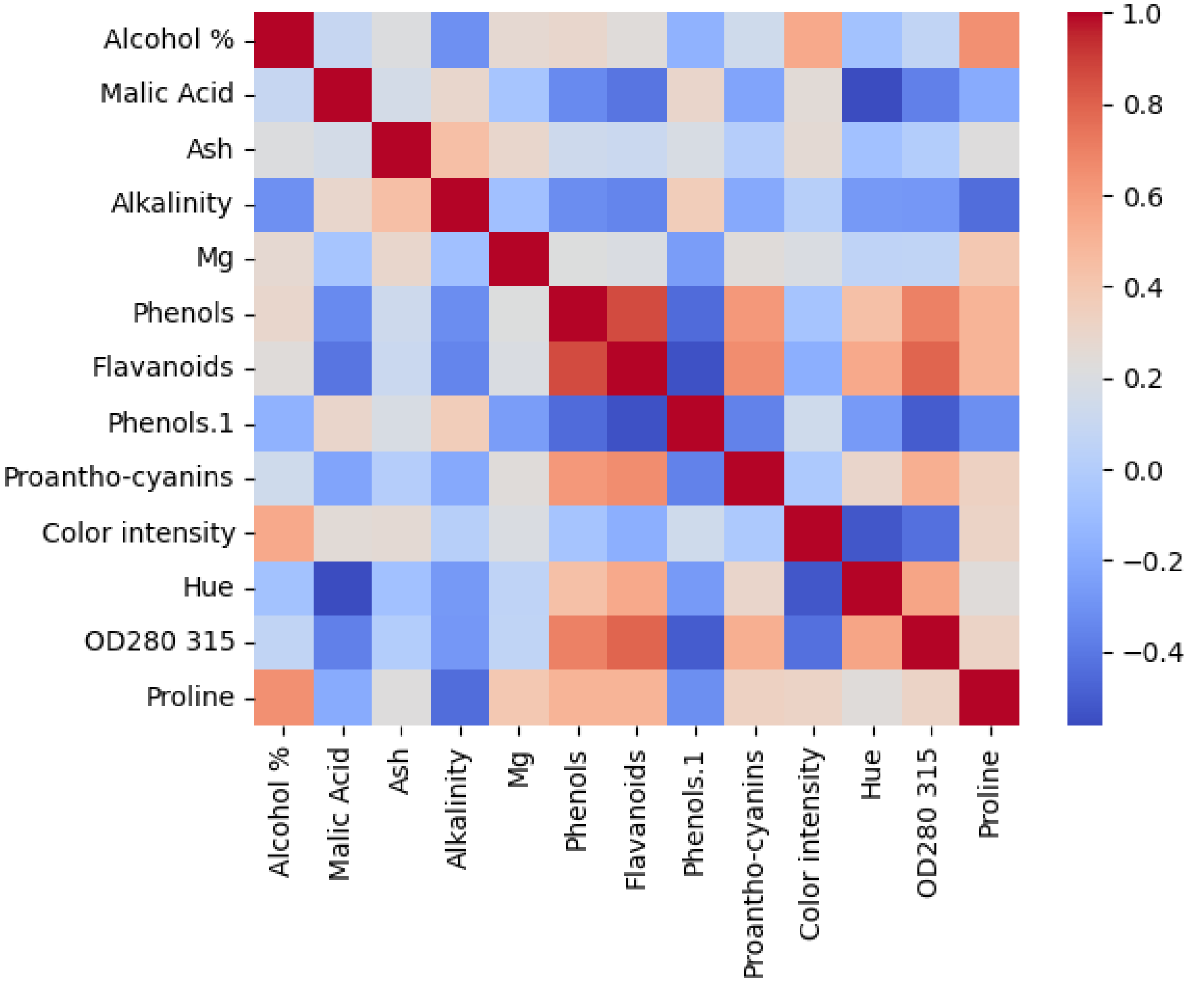
Out[28]:

	Alcohol %	Malic Acid	Ash	Alkalinity	Mg	Phenols	Flavanoids	Phenols.1	Proantho-cyanins	in
0	14.23	1.71	2.43	15.6	127	2.8	3.06	0.28	2.29	
1	13.24	2.59	2.87	21.0	118	2.8	2.69	0.39	1.82	
2	14.83	1.64	2.17	14.0	97	2.8	2.98	0.29	1.98	
3	14.12	1.48	2.32	16.8	95	2.2	2.43	0.26	1.57	
4	13.75	1.73	2.41	16.0	89	2.6	2.76	0.29	1.81	

In [33]:
```
# use numpy to calculate corr coef
# corrmat = np.corrcoef(features)
corrmat = np.corrcoef(features.T)
corrmat.shape
```

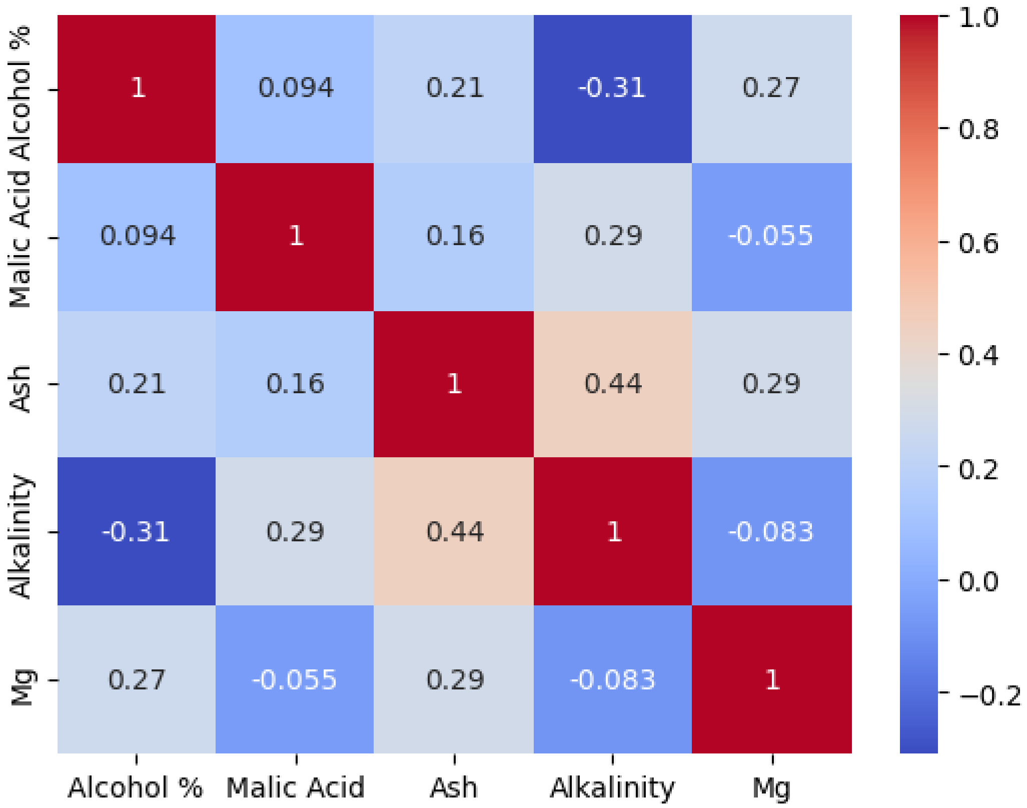
Out[33]:
```
(13, 13)
```

In [36]:
```
# seaborn vis
sns.heatmap(corrmat, cmap='coolwarm', xticklabels=wines.columns, yticklabels
Out[36]: <Axes: >
```



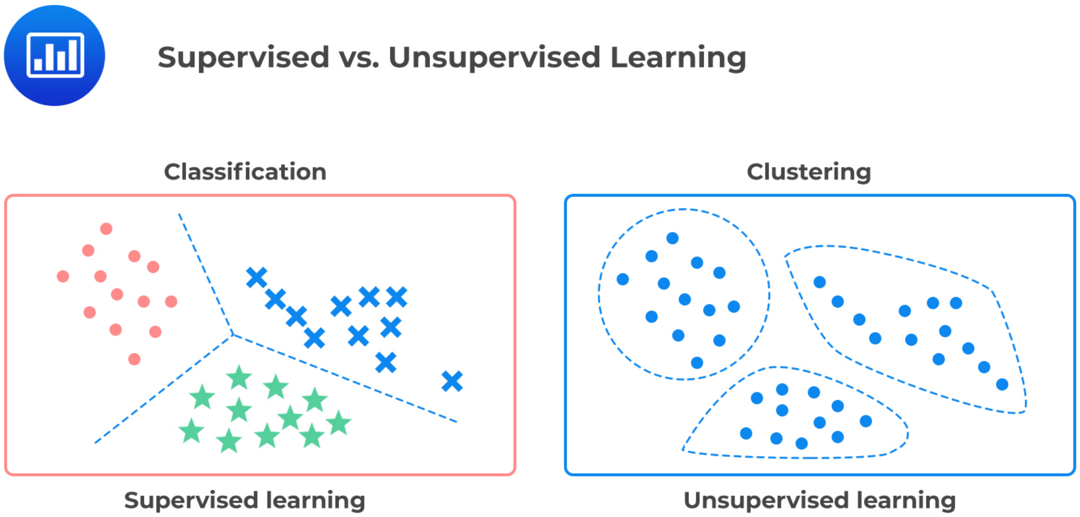
```
In [39]: # seaborn vis
         i = 5
         sns.heatmap(corrmat[:i, :i], cmap='coolwarm', xticklabels=wines.columns[:i],
```

```
Out[39]: <Axes: >
```



Clustering

Clustering is a machine learning technique that involves grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups. It's widely used for exploratory data analysis to find natural groupings, patterns, or structures within data without prior knowledge of group definitions.

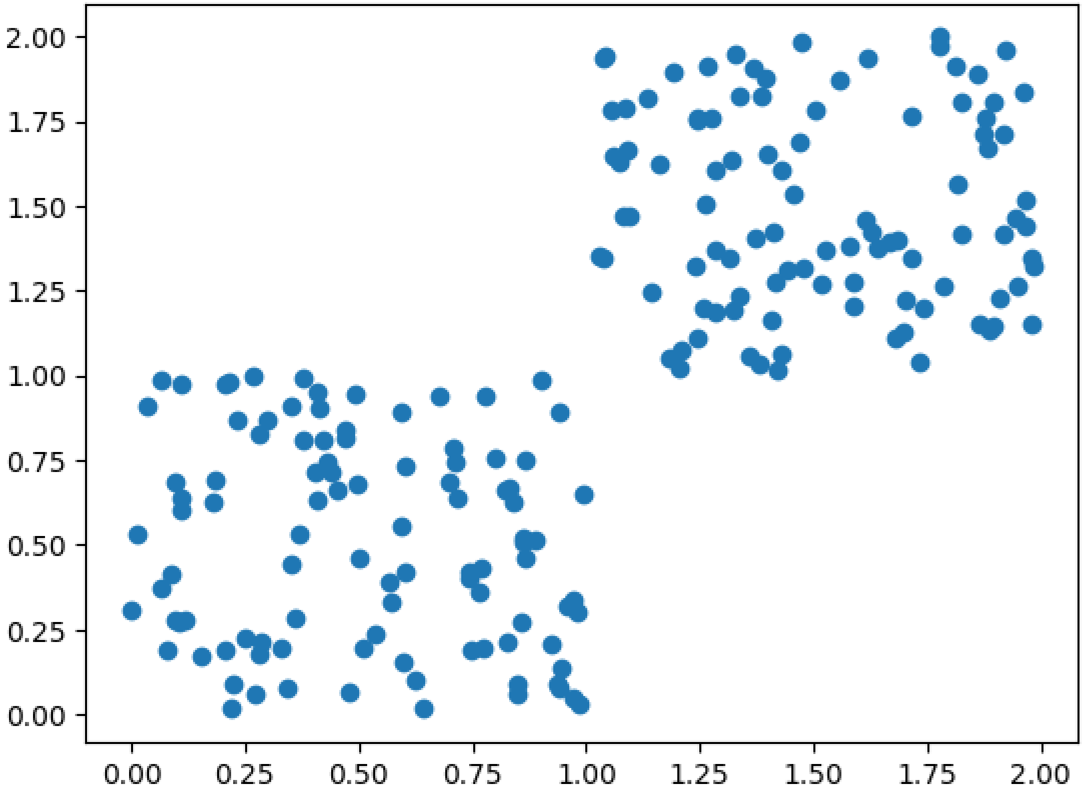


In [40]:

```
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt


def generate_data():
    return np.vstack([
        np.random.random((100, 2)),
        np.random.random((100, 2)) + 1
    ])

data = generate_data()
plt.scatter(data[:, 0], data[:, 1])
```

Out[40]:

```
<matplotlib.collections.PathCollection at 0x29e950580>
```



In [48]:

```
# cluster with K-Means
from sklearn.cluster import KMeans
model = KMeans(4)
model.fit(data)
```

Out[48]:

▾ KMeans

```
KMeans(n_clusters=4)
```

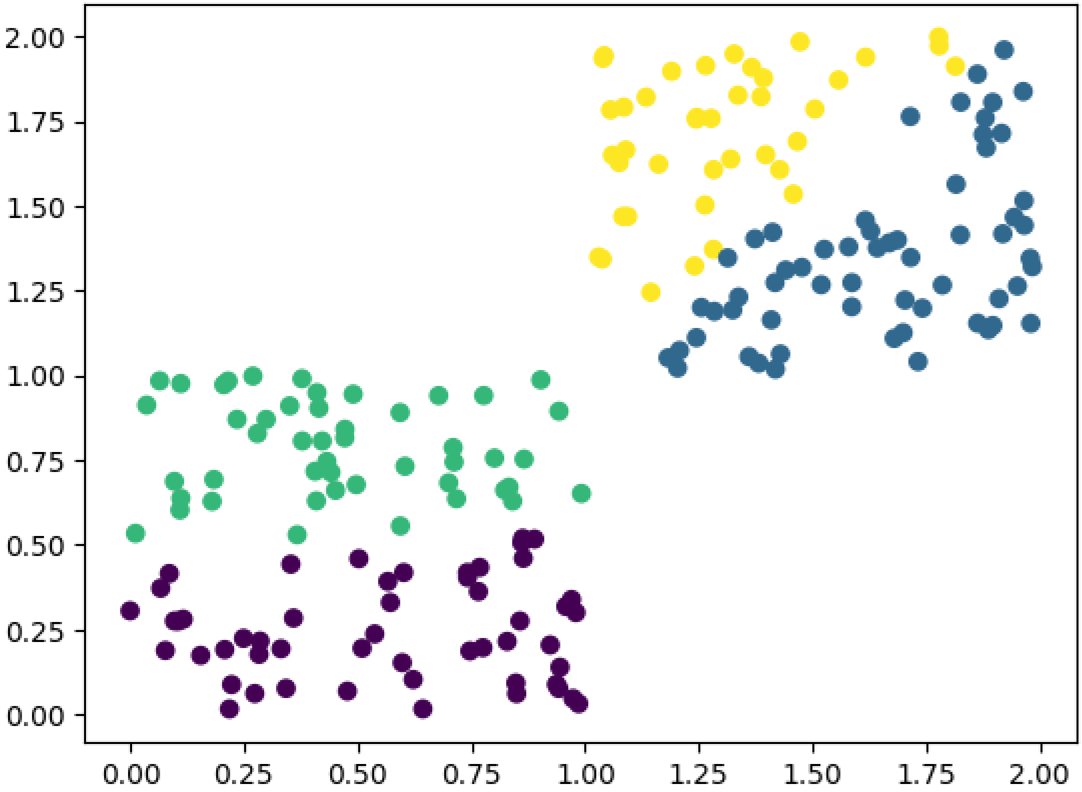
In [49]:

```
model.labels_
```

```
Out[49]: array([2, 0, 0, 0, 2, 2, 2, 2, 0, 2, 2, 0, 0, 0, 2, 2, 0, 0, 2, 0, 2, 0,
                2, 0, 0, 2, 0, 2, 0, 2, 0, 2, 2, 0, 0, 0, 2, 2, 2, 2, 0, 2, 0, 0,
                2, 0, 2, 0, 2, 0, 0, 0, 2, 0, 0, 0, 0, 2, 2, 0, 2, 2, 2, 2, 2, 0,
                0, 2, 0, 0, 2, 2, 2, 0, 2, 0, 0, 2, 0, 0, 2, 0, 0, 0, 2, 0, 2, 0,
                0, 2, 2, 0, 2, 2, 0, 2, 0, 0, 2, 0, 1, 1, 3, 3, 1, 1, 3, 1, 1, 3,
                1, 1, 1, 1, 1, 1, 1, 3, 3, 1, 3, 1, 3, 1, 1, 1, 3, 1, 1, 3, 1, 3,
                1, 3, 1, 1, 3, 3, 1, 1, 3, 1, 3, 3, 1, 1, 1, 1, 1, 1, 3, 3, 3, 1,
                3, 1, 1, 1, 1, 3, 3, 1, 3, 3, 1, 3, 3, 3, 1, 3, 3, 3, 1, 1, 1, 3,
                1, 1, 1, 1, 3, 1, 3, 3, 1, 1, 1, 1, 1, 1, 3, 3, 3, 1, 1, 3, 3, 1,
                1, 1], dtype=int32)
```

```
In [50]: plt.scatter(data[:, 0], data[:,1], c=model.labels_)
```

```
Out[50]: <matplotlib.collections.PathCollection at 0x2a068ab30>
```



Quick Markdown & LaTeX Syntax

Header 1

Header 2

Header 3

List:

- Foo

- Bar

Bold

Italic

Inline Math: $A, B, C, D, \alpha, \beta, \gamma, \lambda, \delta$

Displaymode Math:

$$\frac{\partial f}{\partial X}$$

$$\mathrm{A}, \mathbf{X}$$

Hyperlink

```
In [ ]:
```